Engineering efficient D-xylose fermentation capacity in industrial *Saccharomyces cerevisiae* for advanced bioethanol production

Mekonnen Mezemir Demeke

Supervisors:
Prof. Dr. Johan M Thevelein
Dr. Françoise Dumortier

Members of the Examination Committee:
Prof. Dr. Eckhard Boles
Prof. Dr. Wim Soetaert
Prof. Dr. Alex Verplaetse
Prof. Dr. Patrick Van Dijck
Prof. Dr. Filip Rolland
Dr. Maria Foulquié-Moreno

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Biology

December 2013

Arenberg Doctoraatsschool, W. de Croylaan 6, 3001 Heverlee

ISBN 978-90-8649-681-5
D/ 2013/10.705/96

Acknowledgements

"Life is like riding a bicycle. To keep your balance, you must keep moving." Albert Einstein

My stay and this work have been supported by a predoctoral fellowship from the KU Leuven Interfaculty Council for Development Cooperation (IRO), the EC 7th Framework program (NEMO project) and the Flanders Institute for Biotechnology (VIB).

There are a number of people without which, this work wouldn't have been complete. First of all, my sincere appreciation and thanks goes to my promoter Prof. Dr. Johan Thevelein for giving me the opportunity to work in his lab and excellent support and guidance throughout the work, which contributed a lot to the scientific considerations that I have to adopt. His strong inspiration and trust in me both during the success and failures in the course of this work helped me see the bright future of possibilities in scientific challenges. I would also like to thank Johan for giving me the opportunity to attend a number of scientific meetings in a number of countries, which were important in broadening my scientific understandings and cultural experiences. Thanks to the members of my examination committee for the constructive comments and discussion on the thesis. Special thanks to Prof. Dr. Eckhard Boles for providing us the starting strain used in this study.

I am also very grateful to my co-promoter Dr. Françoise Dumortier for her indispensible help throughout the work. The constructive comments and devotion from Françoise during the experimental work and preparation of this thesis from the beginning to the end was invaluable to the completion of the work. I extend my gratitude to Dr. Maria Foulquié-Moreno for her excellent comments and suggestions both during the experimental designs and preparation of the manuscript. The extensive discussions and collaborations with my colleagues (current and former members of the 'Genetic Analysis' group), and all other members of MCB are greatly appreciated. For that, I thank Yingying, Steve, Tine, Ben, Stijn, Yudi, Thiago Pais, JeanPaul, Annelies, Tom, and all other members of the group; and of course all the staff of MCB.

I thank the very supportive lab technicians who facilitated and assisted in many of the experiments; Paul (currently Master student), Catherina, Evy, Martine, Ilse, Renata, Jan, Thomas, Stephen and Tom Adriany. Thanks Nico for the excellent informatics support whenever needed and important advice during preparations of figures for publications. I extend my gratitude to Dr. Jorge Duitama and Dr. Lieven Clement for the great support during the bioinformatics and statistical analysis. I am very grateful for my master students Tedi, Tung, Mayuko and Tom for their contribution to the project.

My first impression on my first day in the lab was the welcoming and very helpful administrative staff, Hilde and Leni. Their continuous support and amazing arrangements of all kinds of administrative matters whenever necessary was indispensible. Thank you very much Hilde and Leni. Thank you to all the staff of MCB for the very nice working environment and various social activities in the lab.

The work in the lab cannot be separated from the life outside the lab. I was very blessed to have a loving family who were there for me in time of difficulties and sharing my happiness. Thank you Nardi for the wonderful love, care and prayer, as well as understanding when I was obsessed with

the ‘experiment’ stuff. Thank you princess Lael for making my life sweeter and giving me a happy ‘time out’ moments from the obsession. During the great challenge we encountered in the family, the contribution of my mother in law Laku, and bro Gashu was indispensible. My deep appreciation extends to my parents and to my brother Mengistu and my sister in law Zimam, for their sacrificial support to see me succeed in my carrier from my child hood to the end of my study.

Thank you to all my Ethiopian friends in Belgium, our church choir team and all church members for the wonderful time together. Special thanks to Meron, for sharing our PhD challenges, your great support, advice and of course the help on the data sorting and other computer issues. I don’t forget the wonderful stress releasing times we share together, especially in ALMA, with Meron, Mehari, Tekle, and other former and current PhD students; thank you all.

Last and most, thanks to the Almighty God for giving me all these pleasant people around me, and keeping me in balance in riding the bicycle of life.

Summary

Bioethanol production using non-food biomass such as agricultural wastes, energy crops, and forest residues as substrates, is advancing to commercial production scale. However, economically competitive ethanol production from such lignocellulosic biomass remains a great challenge. One of the reasons is the lack of microorganisms that efficiently convert all the sugars present in the lignocellulose hydrolysates into ethanol under industrial conditions.

The baker's yeast *Saccharomyces cerevisiae* is still the dominant organism for industrial ethanol production owing to its high rate of fermentation of hexose sugars, very good tolerance to ethanol and to inhibitors in lignocellulosic hydrolysates. Unfortunately, baker's yeast is unable to metabolize pentose sugars, particularly D-xylose, which accounts for up to 35% of total sugars in lignocellulosic biomass. Efficient utilization of D-xylose, in addition to inhibitor tolerance, is required in organisms to be used for cost effective and sustainable production of ethanol from lignocellulosic material.

Our general objective was therefore to construct a robust yeast strain that efficiently converts D-xylose into ethanol and is tolerant to inhibitors present in lignocellulosic hydrolysates. For that purpose, this project started with one of the most widely used first generation bioethanol production yeast strains (Ethanol Red), in which a *Clostridium phytofermentans xylA* based D-xylose and an L-arabinose gene cassette had previously been inserted. Despite the presence of all known genes required for D-xylose and L-arabinose utilization in the genome, the recombinant industrial strain was unable to utilize any D-xylose or L-arabinose. We applied a systematic evolutionary engineering approach (random mutagenesis, genome shuffling followed by selection in a D-xylose-enriched lignocellulose hydrolysate, and adaptive evolution in D-xylose) to establish rapid D-xylose fermentation capacity in the recombinant strain. Consequently, we were able to develop a robust D-xylose fermenting strain with a productivity and yield of ethanol that was higher than that of any reported recombinant industrial strain of *S. cerevisiae.* The evolved strain GS1.11-26 demonstrated substantial tolerance to inhibitor-containing lignocellulosic hydrolysates producing ethanol with a yield close to the maximum theoretical yield. However, the evolved strain GS1.11-26 showed a partial respiratory defect causing a reduced aerobic growth rate and it also had a slightly reduced inhibitor tolerance compared to the original strain.

To enhance the direct industrial applicability of the strain, we further eliminated the negative properties of GS1.11-26 (slow growth rate, and reduced inhibitor tolerance) using a meiotic recombination tool. The method combines the superior genetic elements from

the genomes of the D-xylose utilizing strain GS1.11-26 with either a haploid segregant of Ethanol Red or a highly inhibitor tolerant diploid strain Fseg25. In this way, we successfully developed three robust industrial yeast strains that combine efficient D-xylose utilization with high inhibitor tolerance for use in bioethanol production with lignocellulose hydrolysates. From the economic point of view, the use of these strains will significantly contribute to the reduction of the overall ethanol production cost, especially from second generation feedstocks.

Using whole genome sequence comparison between the evolved and the original parent strain, amplification of the *xylA* gene has been identified and experimentally verified as a crucial but not the only genetic change responsible for the high D-xylose utilization rate in GS1.11-26. In addition, we showed that an extra-chromosomal circular DNA (eccDNA) carrying *xylA* has been formed in the process of evolutionary adaptation. Chromosomal integration of the eccDNA over the course of the evolutionary adaptation steps as tandem repeat resulted in amplification of *xylA* and was stable for several generations without selection. The high copy number of *xylA* correlated with the high activity of D-xylose isomerase in GS1.11-26. Quantitative trait loci (QTL) mapping using a modified pooled segregant whole genome sequence analysis resulted in the identification of at least three genomic loci that are linked to the fast D-xylose fermentation rate in GS1.11-26. After evaluation of one of the QTLs by reciprocal hemizygosity analysis, a mutation in a gene that hadn't been associated previously with D-xylose fermentation was found to be linked with the high D-xylose utilization rate in GS1.11-26.

The genetic changes identified in this study, together with other genetic factors yet to be identified in the other QTLs may be transferred to other target industrial strains to endow efficient D-xylose fermentation capacity. We also showed for the first time that QTL mapping can be performed directly in diploid or aneuploid strains without the need to isolate haploid derivatives. This will facilitate the polygenic analysis of important industrial traits in aneuploid or polyploid industrial strains.

Samenvatting

De productie op commerciële schaal van bio-ethanol op basis van biomassa die niet concurrentieel is met de voedselindustrie, zoals land- en tuinbouw afval, energiegewassen of bosbouwresten, komt stilaan op gang. Niettemin blijft een economisch haalbare productie op basis van lignocellulose biomassa een grote uitdaging. Eén van de struikelblokken is het ontbreken van microorganismen die op industriële schaal alle suiker aanwezig in een lignocellulose hydrolysaat efficiënt in ethanol kunnen omzetten.

De bakkersgist *Saccharomyces cerevisiae* blijft het belangrijkste microorganisme in de industriële bio-ethanol productie vanwege zijn snelle vergisting van hexose suikers en zijn hoge tolerantie voor ethanol en inhibitoren die in lignocellulose hydrolysaten aanwezig zijn. Helaas kan bakkersgist pentose suikers, en in het bijzonder D-xylose (dat tot 35% van de totale suikerconcentratie in een lignocellulose hydrolysaat kan uitmaken) niet op een efficiënte wijze metaboliseren. Naast efficiënte fermentatie van D-xylose moet een microorganisme tevens een hoge tolerantie hebben t.o.v. inhibitoren aanwezig in lignocellulose hydrolysaten om op een kost-effectieve en duurzame manier bio-ethanol uit lignocellulose hydrolysaten te kunnen produceren.

Onze algemene doelstelling was dan ook een robuuste giststam te maken die D-xylose, aanwezig in inhibitor-rijke hydrolysaten efficiënt kan omzetten naar ethanol. We zijn vertrokken van Ethanol Red, één van de meest gebruikte eerste generatie bio-ethanol productiestammen. Eerder werd in die stam een xylose cassette, gebaseerd op het *xylA* gen van *Clostridium phytofermentans* ingebracht evenals een arabinose cassette. Niettegenstaande alle gekende genen die nodig zijn voor het metaboliseren van D-xylose en arabinose in de stam aanwezig waren, was deze niet in staat D-xylose of L-arabinose te gebruiken. Via systematische “evolutionary engineering” (random mutagenese, “genome shuffling” gevolgd door selectie in D-xylose verrijkt lignocellulose hydrolysaat en adaptieve evolutie in D-xylose) beoogden we van de recombinante stam een efficiënte D-xylose fermenterende stam te maken. Uiteindelijk zijn we erin geslaagd een robuuste D-xylose fermenterende stam te maken met een ethanol productie en opbrengst die tot nu toe bij geen enkele recombinante industriële *S. cerevisiae* stam gerapporteerd werd. De bekomen stam, GS1.11-26, vertoont een substantiële tolerantie in inhibitor-rijk lignocellulose hydrolysaat en geeft een ethanol opbrengst die de maximale theoretische opbrengst benadert. De stam vertoonde niettemin een partieel respiratorisch defect dat resulteerde in een verminderde aerobe groeisnelheid; verder werd er ook een licht verminderde inhibitor resistentie vastgesteld.

Om GS1.11-26 te kunnen gebruiken voor de industriële ethanol productie was het nodig die verder te verbeteren (vooral de groeisnelheid en inhibitor tolerantie). Hiervoor werd een nieuwe methode van meiotische recombinatie toegepast waardoor er een combinatie

ontstond van de superieure genetische elementen uit de genomen van de D-xylose fermenterende GS1.11-26 stam en ofwel een Ethanol Red segregant (die een goede aerobe groei vertoont), en ofwel een hoge inhibitor tolerante diploide stam Fseg25. Op die manier werden met succes drie inhibitor tolerante stammen ontwikkeld die D-xylose in lignocellulose hydrolysaten efficiënt fermenteren. Vanuit economisch oogpunt zal het gebruik van deze stammen de productiekosten van tweede generatie bio-ethanol substantieel verlagen.

Sequenering van het ganse genoom van de originele *xylA* bevattende stam en van de geëvolueerde stam toonde aan dat amplificatie van het *xylA* gen een cruciale rol speelt in de verbetering van de stam voor D-xylose fermentatie, maar daarnaast waren ook nog andere genetische veranderingen nodig. Verder hebben we ook aangetoond dat er gedurende het proces van evolutionaire adaptatie een extra-chromosomaal circulair DNA (eccDNA) gevormd werd, dat later als een “tandem-repeat” terug geïntegreerd werd. Deze amplificatie bleek stabiel over verschillende generaties zonder de selectiedruk van D-xylose, en was gecorreleerd met een hoge D-xylose isomerase activiteit. Door middel van sequenering van het volledige genoom van een pool van fenotypisch geselecteerde segreganten konden minstens drie loci gevonden worden die gekoppeld zijn aan de snelle xylose fermentatie van GS1.11-26. Bij het analyseren van één van die QTLs (Quantitative trait loci) via reciproke hemizygosity analyse werd een mutatie gevonden in een gen dat voordien nooit in verband werd gebracht met D-xylose; deze mutatie blijkt geassocieerd te zijn met de snelle fermentatie capaciteit van GS1.11-26.

Transfer van de genetische veranderingen die in deze studie gevonden werden, samen met nog een aantal te identificeren genetische factoren uit de andere QTLs, naar industriële doelstammen zal leiden tot efficiënte D-xylose fermenterende productiestammen. We hebben tevens voor het eerst aangetoond dat het mogelijk is QTL kartering uit te voeren met diploïde of aneuploïde stammen zonder dat het nodig is eerst naar haploïde afgeleiden over te gaan. Zodoende wordt polygenische analyse van belangrijke industriële eigenschappen in aneuploïde en polyploïde industriële stammen veel gemakkelijker.

List of abbreviations

Chr	Chromosome
CNV	copy number variation
DNA	Deoxyribonucleic acid
DW	dry weight
eccDNA	extra-chromosomal circular DNA
EMS	ethyl methanesulfonate
HPLC	high performance liquid chromatography
LB	Luria-Bertani
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
noPPP	non oxidative pentose phosphate pathway
OD	optical density
PCR	polymerase chain reaction
PPP	pentose phosphate pathway
PSS	pooled segregant whole genome sequence analysis
QTL	quantitative trait locus
RKI	ribose-5-phosphate ketol-isomerase
RNA	ribonucleic acid
RPE	D-ribulose-5-phosphate 3-epimerase
SC	synthetic complete medium
SHF	separate hydrolysis and fermentation
SNP	single nucleotide polymorphism
SSF	simultaneous saccharification and fermentation
TAL	transaldolase
TKL	transketolase
tRNA	transfer RNA
VHG	very high gravity
XDH	xylitol dehydrogenase
XI	D-xylose isomerase
XK	xylulokinase
XR	D-xylose reductase
YPD	yeast extract peptone + Dextrose
YPX	yeast extract peptone + D-xylose

Table of Contents

General Introduction and objectives

Bio-ethanol is currently the dominant renewable biofuel used in the transportation sector. It has already been introduced on a large scale in various countries, such as Brazil, the US, and increasingly in European countries. Currently, bioethanol is predominantly produced from food crops such as corn, wheat, sugar beet and sugarcane. Such starch or sucrose based ethanol production is known as 'first generation' bioethanol production. However, the use of food crops for fuel production remains controversial. In the past few years, substantial efforts have been focused on production of bioethanol from non-food biomass such as agricultural and forest residues, energy crops, and waste streams. Such feedstocks are composed of cellulose, hemicellulose and lignin (collectively termed lignocelluloses). Cellulose is a polymer of glucose, while hemicellulose is a complex polysaccharide that constitutes hexose and pentose sugars. Bio-ethanol production from such lignocellulosic biomass, also known as 'advanced' bioethanol production, is considered to be one of the most sustainable and environmental-friendly alternative fuel productions for the transport sector. After decades of research on improvement of bioethanol production from such biomass, economically viable production remains a great challenge. Lack of yeast strains that ferment all the sugars in the lignocellulosic biomass under industrial condition is among the major obstacles in the commercialization of bioethanol production from lignocellulosic biomass.

The yeast *Saccharomyces cerevisiae* is the best organism for industrial ethanol production owing to its high rate of fermentation of hexose sugars, high tolerance to ethanol, inhibitors, acidity and other industrial process conditions, well-established production, storage and transport systems at commercial scale, comprehensive physiological and molecular knowledge, and its genetic tractability. However, it is unable to metabolize pentose sugars, particularly D-xylose, which represent up to 35% of total sugars in lignocellulosic feedstocks. Thus, efficient utilization of D-xylose is required for cost effective and sustainable production of ethanol from lignocellulosic material. In addition, pretreatment and hydrolysis of lignocellulosic biomass to release free sugars result in the release of not only the fermentable sugars but also various inhibitors that hampers the growth of yeasts, making extreme inhibitor tolerance a crucial trait for reaching economically viable advanced bioethanol production. The inherently higher robustness and tolerance of *S. cerevisiae* to various inhibitors gives it a head start in programs aimed at developing strains with extreme inhibitor tolerance, able to efficiently ferment hexoses and pentoses (mainly D-xylose) in concentrated lignocellulose hydrolysates.

Considerable progress has been made in the past few years in the development of yeast strains that are able to ferment D-xylose through heterologous expression of D-xylose metabolism pathways from natural D-xylose utilizing yeasts or bacteria into *S. cerevisiae*. Improving natural D-xylose utilizing yeast species and bacteria for inhibitor tolerance has also been attempted.

However, those strains do not meet the industrial standards, in particular the demand for high yield and productivity in non-detoxified lignocellulosic hydrolysates. Since pentose fermentation appears to be much more sensitive to the toxic inhibitors, the productivity of the yeast in high-density lignocellulose hydrolysates is largely determined by the strain's robustness during pentose fermentation.

The goal of this study was to construct such a robust D-xylose utilizing industrial *S. cerevisiae* strain that is also tolerant to various inhibitors present in a variety of lignocellulosic biomass hydrolysates. This thesis starts with literature overview (chapter 1) on the current developments in advanced bioethanol production technologies with more emphasis on engineering yeast strains for D-xylose utilization. Chapters 2 through chapter 6 explain the results obtained from this study. Chapter 2 illustrates the development of a robust D-xylose utilizing strain GS1.11-6 using a combination of random mutagenesis, genome shuffling and evolutionary adaptation in D-xylose medium. The evaluation of the strain both in laboratory medium and in three industrially relevant lignocellulosic feedstocks is reported in chapter 3. Due to the negative effect of the mutagenesis on GS1.11-26, further upgrading of the strain has been done to eliminate those negative properties and to further improve the inhibitor tolerance of the D-xylose utilizing strain using meiotic recombination, which is described in chapter 4. The last two chapters elucidate the identification of some of the genetic basis behind high D-xylose utilization capacity in the GS1.11-26 using whole genome sequence comparison (chapter 5), and pooled segregant whole genome sequence analysis (chapter 6). Finally, the main results obtained in this study is summarized and with some recommendations in section 'General conclusion and recommendations'.

Chapter 1:
Literature review

1. Literature review

1.1. Introduction to biofuels

Global concern for increased green house gas emissions due to fossil fuel burning, together with the depletion of oil reserves and geopolitical issues, drive the search for alternative, environmentally friendly energy sources. These alternative energy supplies should simultaneously acquire sustainability and substantial reduction of green house gas emission.

More than half of the green house gas emission is attributed to power generation and the transport sector. To date, most of the alternative energy options are suited for power generation. While solar and wind energy are among the promising future green energy sources for power generation, the increased use of biofuels such as bioethanol in the transportation sector is considered to be among the strongest energy efficiency gains (Goldemberg, 2007). The use of liquid *biofuel*s for the transportation sector has significant importance in reducing the dependency on petroleum oil. It also generates less green house gases than petroleum derived fuels provided that efficient methods of production are developed (Schwietzke *et al.,* 2008).

Based on the maturity of the production technology, bioethanol production can be classified as "conventional" or "advanced".

1.1.1. Conventional biofuel technology:

The production of conventional biofuels, also known as first generation biofuels, uses well established technologies that are already at commercial scale. The two best examples are bioethanol produced from crops such as corn and sugarcane, and biodiesel produced from oil crops such as soy bean, sunflower and rapeseed oil. Blending of bioethanol with gasoline and biodiesel with diesel oil can significantly contribute to green house gas emission savings (Randelli, 2009).

Bioethanol, in particular, is considered to be one of the most promising and feasible short-term to medium-term alternative transportation fuels. In the year 2011 alone about 100 billion liter of biofuels were produced and bioethanol accounts for more than 80%. Corn ethanol in the US and sugarcane ethanol in Brazil share about 88% of the global output. Moreover, estimates based on investment plans suggest that bioethanol production worldwide will increase to 100 billion liters in the year 2014 (Sims *et al.,* 2010).

However, production of ethanol and other biofuels from food crops remains under close scrutiny due to a number of reasons, including the impact on food price and shortage, its meager contribution to reducing overall greenhouse gas emission and the high cost of production, which makes it less competitive with fossil fuels (Fairley, 2011). Therefore, the

conventional biofuels do not seem to be a good solution for the rising energy demand and greenhouse gas emission problem. On the other hand, development of advanced biofuels can be a major part of the answer.

1.1.2. Advanced biofuel technology

Biofuels produced from non-food lignocellulosic biomass such as agricultural and forest residues, municipal solid wastes and energy crops are believed to be the future solution to the energy deficit and the problem with green house gas emissions (Randelli, 2009). Such lignocellulosic materials constitute the most abundant organic materials in the biosphere and are considered to be the best future substrates for *biofuel* production (Nevoigt, 2008). Although, significant technological developments have been achieved in the field in the last three decades, advanced biofuel production is not yet economically feasible, and is currently only at the pilot or demonstration stage (Macrelli *et al.,* 2012; Mielenz, 2001; Sims *et al.,* 2010)

Bioethanol is currently the dominant renewable biofuel that is being used in the transportation sector (Sánchez and Cardona, 2008). It has already been introduced in large scale in various countries such as Brazil, the US and European countries and is now predominantly produced from food crops. In the past few years, considerable efforts have been focused on production of ethanol from non-food biomass. The potential of ethanol produced from lignocellulosic wastes such as crop residues and sugar cane bagasse has been shown to contribute significantly to the overall replacement of the fossil fuel gasoline (Kim and Dale, 2004).

A major challenge in advanced bioethanol production is reducing the relatively high cost of production. This challenge is primarily attributed to the lack of efficient and cheap technologies to release the fermentable sugars from the lignocellulosic feedstock, and the lack of robust microorganisms that can ferment all the sugars present in the biomass hydrolysate efficiently (Schubert, 2006). Only when these challenges will have been overcome, the potential of advanced biofuels can be realized.

1.1.3. Bioethanol production process

The key steps in the bioethanol production process, in general, comprise size reduction of the biomass, pretreatment, hydrolysis, fermentation and distillation (Figure 1.1). In conventional bioethanol industries that use starch or sugar based biomass, the pretreatment step in not required or is minimal. In sucrose-based ethanol industries, which are dominant in Brazil (sugar cane) and Europe (sugar beet), the fermentable sugar sucrose is directly obtained in the juice, which makes the use of enzymes unnecessary. In starch-based industries, the grains are first milled into course flour. The flour is then mixed with water, and enzymes (amylase and glucoamylase) are added to hydrolyze starch into glucose. The hydrolyzed slurry, called mash,

containing the fermentable sugars, is then fermented by microorganisms (predominantly yeast) for 2 to 3 days in a fermentation tank. Ethanol is separated in the distillation tank. After the distillation step, the ethanol still contains about 5% water. Further purification is done by passing through a molecular sieve, which generates high grade ethanol that can be used directly as a fuel. The stillage that remains in the distillation tank (containing solids from the grains, water and yeast) is processed to produce 'dried distillers grain with soluble' (DDGS) or 'wet distillers grain' (WDG) which is generally used for cattle feed (Schwietzke *et al.,* 2009).

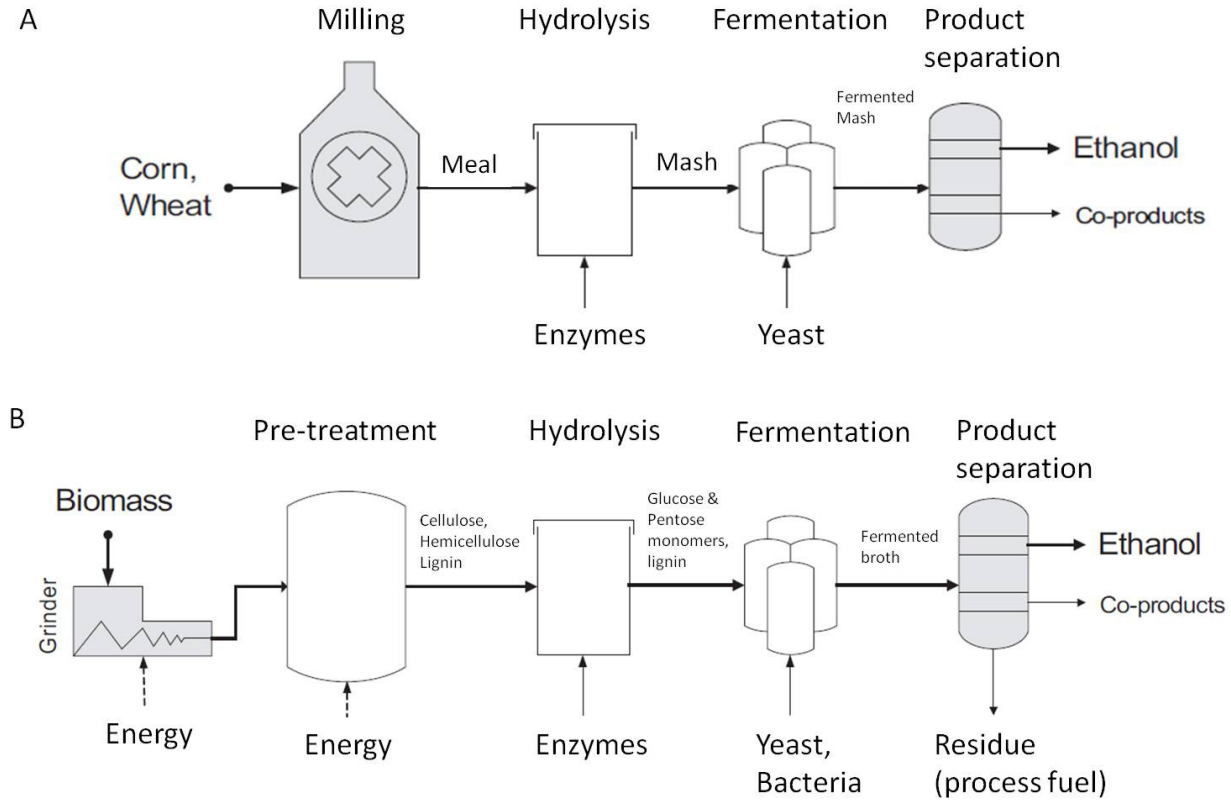


Figure 1.1. Bioethanol production process. (A) Conventional ethanol production from starch based biomass. The first milling step is used to ground the biomass into course mill. The meal is hydrolyzed using starch degrading enzymes into glucose, which is then fermented into ethanol by natural yeast. The ethanol is separated by distillation, and a further dehydration step results in pure ethanol. (B) Advanced bioethanol production from lignocellulosic biomass is generally similar but includes additional pretreatment before hydrolysis to make the cellulose and hemicelluloses accessible to enzymatic hydrolysis (Schwietzke *et al.,* 2009).

In contrast to the conventional bioethanol production process, the biomass for advanced (second generation) bioethanol industries requires a harsh pretreatment step before enzymatic hydrolysis. This is due to the complex composition of the woody lignocellulosic biomass, where cellulose and hemicellulose are not easily accessible for enzyme action. This pretreatment step, which requires energy, is one of the reasons for the high cost of production of advanced bioethanol from such woody biomass (discussed below).

1.2. Biomass for advanced bioethanol production

1.2.1. Composition

As described above, advanced biofuels are made from plant biomass that does not compete with food provision. Such biomasses are widespread and constitute the most abundant sources of fermentable sugars. These include agricultural by-products (such as cereal straw and sugar cane bagasse), forest residues, organic municipal solid wastes, dedicated feedstocks (such as switchgrass, giant reed) and short rotation woody crops (Hoogwijk *et al.,* 2003; Sims *et al.,* 2010). Such biomass materials are chiefly composed of cellulose, hemicellulose and lignin, therefore called lignocelluloses (Figure 1.2). They also contain small amounts of pectin, protein, extractives and ash. Cellulose is a polymer of D-glucose; while hemicelluloses are more complex heterogeneous polysaccharides made of both hexose and pentose sugars. Although the proportion of the various components varies significantly with the type of biomass, the dry weight of lignocellulosic biomass in general contains up to 75% carbohydrates, mainly D-glucose, and a significant proportion of pentose sugars, predominantly D-xylose and some L-arabinose (Schadel *et al*, 2010). The composition of various lignocellulosic biomasses is shown in Table 1.

Table 1.1 Composition of different lignocellulosic biomass materials (adapted from Jørgensen *et al.,* 2007). Glucose is mainly derived from cellulose, while the remaining sugars are mainly derived from the hemicellulose fraction.

Material	Glucose	D-xylose	L-Arabinose	Mannose	Lignin
	(% total dry weight)				
Hard wood					
Birch	38.2	18.5	- [a]	1.2	22.8
willow	43.0	24.9	1.2	3.2	24.2
Soft wood					
Spruce	43.4	4.9	1.1	12.0	28.1
Pine	46.4	8.8	2.4	11.7	29.4
Grasses/poaceae					
Wheat straw	38.2	21.2	2.5	0.3	23.4
Rice straw	34.2	24.5	ND[b]	ND[b]	11.9
Corn stover	35.6	18.9	2.9	0.3	12.3
Bagasse	39.0	22.1	2.1	0.4	23.1
Switch grass	31.0	0.4	2.8	0.2	17.6
Giant reed	39.3	18.4	1.2	0.2	26.2

[a]Below detection limit; [b]ND, not determined

Cellulose, the most abundant polymer on earth, is a linear polymer of D-glucose linked by beta-1,4-glucosidic bonds. The number of glucan units in the polymer, called the degree of polymerization, ranges from 2000 to 27,000, depending on the type of plant (Taherzadeh, 2007). Cellulose exists as a crystalline structure in the plant cell wall, with intramolecular and

intermolecular hydrogen bonds, making it resistant to chemical, physical and enzymatic degradation.

Hemicellulose has a lower degree of polymerization (150 to 200) compared to cellulose. It is composed of complex polysaccharide units with linear or branched units of sugars such as D-glucose, D-xylose, D-galactose, D-mannose and L-arabinose. Hemicelluloses account for 11-37% of the total lignocellulose component, and the percentage of the different monomeric sugars in hemicelluloses varies with the type of biomass. While soft wood hemicelluloses contain more hexose sugars than hard woods, the pentose sugar D-xylose accounts for the highest proportion of sugars in hemicelluloses, especially in hard woods (Palmqvist and Hahn-Hägerdal, 2000).

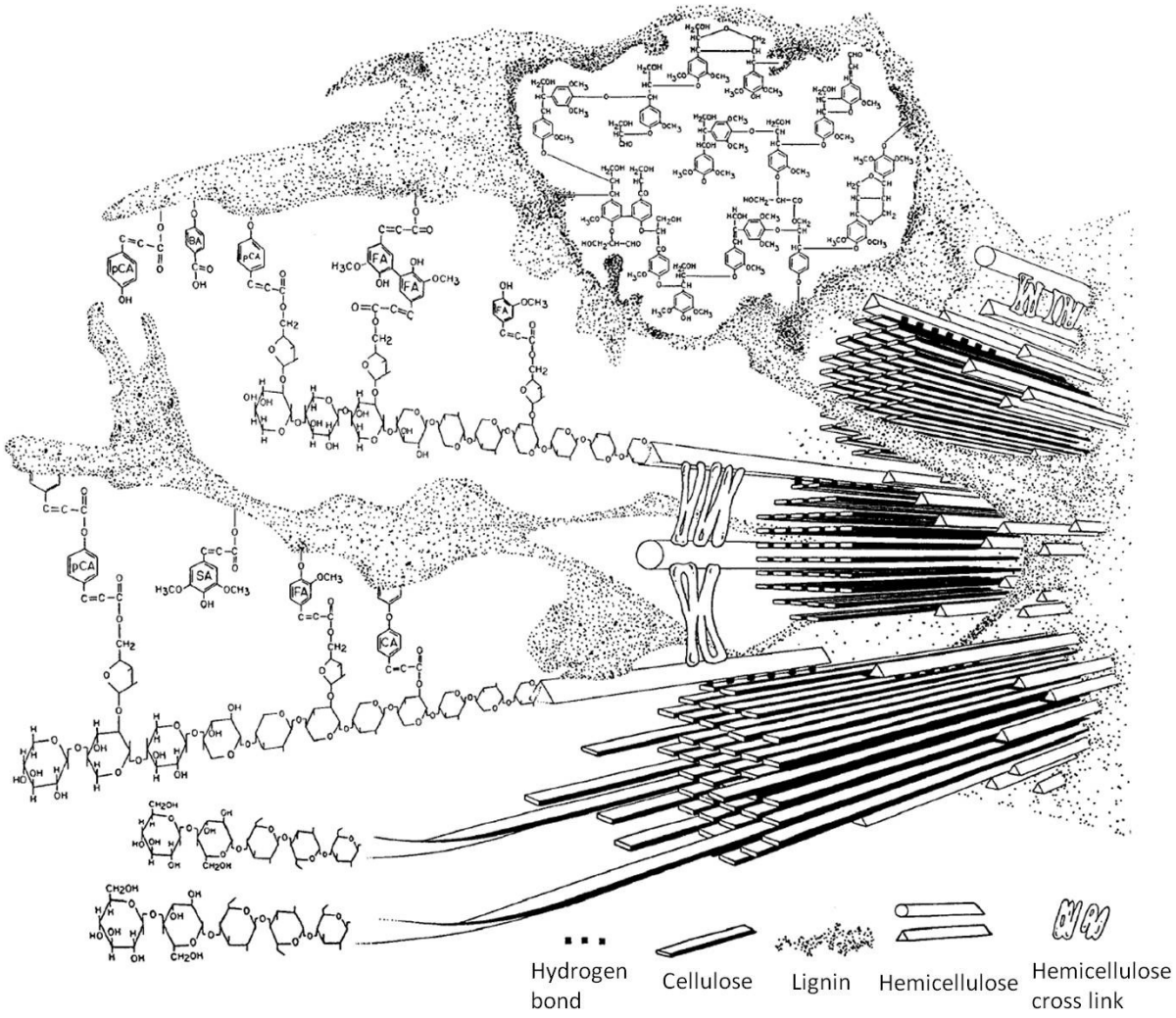


Figure 1.2. Structure of the secondary cell wall (Bidlack *et al.,* 1992), showing the complex organization of the different components of the cell wall. The rigid crystalline structure of cellulose is shown as a long ribbon-like element that is held together by hydrogen bonds. The hemicellulose structure is represented by long triangular or rod shaped cylinders. Lignin is shown as an embedding matrix of polymerized lignin monomers. Various branched chains of hemicellulose and ester bonds between lignin and hemicellulose are also indicated.

Lignin is another structural component of plant cell wall. It is a very complex aromatic compound primarily made of polymers of phenylpropane units. It forms a chemical bond with cellulose and hemicelluloses providing the cell wall with extreme resistance to chemical and enzymatic degradation. It is the most abundant non-carbohydrate fraction of plant cell wall. The proportion of lignin is higher in soft woods than hard woods (Taherzadeh and Karimi, 2007).

In general, the three polymers; cellulose, hemicelluloses and lignin provide the cell wall and thus, the whole plant with a strong rigidity. The crystalline structure of cellulose itself is highly stable. This structure is further attached to the hemicelluloses, which forms a variety of inter- and intramolecular bonds. Lignin glues both cellulose and hemicelluloses together making the plant cell wall very difficult to be broken down. Due to this complexity of the structure of lignocellulosic biomass, hydrolysis of the sugar polymers into fermentable sugars is very challenging and requires a harsh pre-treatment step before the enzymatic hydrolysis step (Sanderson, 2011).

1.2.2. Pre-treatment of lignocellulosic biomass

The general aim of pretreatment is to enhance the accessibility of the cellulose to enzymes through solubilizing hemicellulose and lignin, and decreasing the crystallinity of cellulose (Figure 1.3). It may also include removal of lignin and hemicellulose or alteration of their structure. An ideal pretreatment mechanism should provide maximum accessibility of the polysaccharides to enzymes, with minimum formation of degradation products that inhibit both hydrolyzing enzyme and fermenting yeast and with minimum loss of sugars. Various methods of pretreatment exist currently and the application of the different technologies depends on the type of biomass and its composition. The methods are mainly classified as physical and chemical processes but also are applied as a combination of both (Chandra *et al.,* 2007).

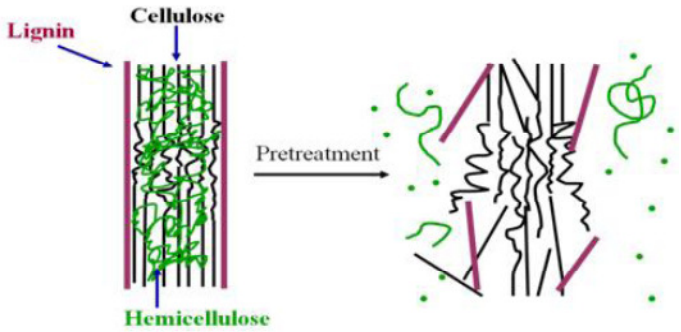


Figure 1.3. Effect of pretreatment on the structure of lignocellulosic biomass (Mosier *et al.,* 2005). The pretreatment disrupts both the lignin seal and the crystalline structure of cellulose. Some monomeric sugars (green dots) are also released by pretreatment mainly from the hemicellulose fraction.

1.2.2.1. Milling

Physical pretreatment such as milling acts by breaking down the biomass into smaller particles. It reduces the crystallinity and degree of polymerization of cellulose making it more accessible to enzymatic hydrolysis. Though this method can be effectively used on a wide range of biomass, the technique is relatively energy intensive. Moreover, such methods do not remove the lignin fraction that is known to hinder access of enzymes to cellulose and directly inhibit the activity of enzymes (Taherzadeh and Karimi, 2008). Although the method of mechanically reducing biomass particles can sometimes be used for easier material handling in the subsequent processing steps, it is not commonly applied as sole means of pretreatment.

1.2.2.2. Steam explosion

Steam explosion is one of the most efficient and commonly used methods of pretreatment that has been tested with various types of biomass in both laboratory and pilot scale experiments. It employs high pressure between 0.7 and 4.8 MPa and high temperature ranging from 160-260°C for several seconds or up to 20 minutes. The pressure is then suddenly dropped to allow explosive decompression of the biomass. In this method, the hemicellulose fraction is usually hydrolysed in the liquid phase partly by the organic acids that are released from the biomass and partly by the acidic nature of water when heated to high temperature. In general the major effect of steam explosion is the removal of hemicelluloses and the transformation of lignin, which increases the accessibility of cellulose to cellulases. The main disadvantage of this method is the low yield of sugars from the hemicellulose fraction due to its degradation, and the formation of inhibitory compounds (Zheng *et al.,* 2009).

1.2.2.3. Catalyzed steam explosion

Addition of catalysts such as H_2SO_4 to the biomass prior to steam explosion is more efficient than steam explosion alone. It has been shown that impregnation of biomass with H_2SO_4 increased the overall yield of fermentable sugars by enhancing the solubilization of hemicelluloses (Ballesteros *et al.,* 2006). The use of SO_2 instead of H_2SO_4 has been proven to be more effective since it produces more digestible substrates and releases more D-xylose from the hemicellulose fraction. Catalyzed steam explosion is among the best pretreatment options not only due to its effectiveness but also because of the energy savings compared to regular steam exploision (requires lower energy input). However, a number of inhibitors are released due to the degradation of carbohydrates. In addition, some part of the xylan fraction is lost and the lignin-carbohydrate matrix is not completely disrupted (Mackie *et al.,* 1985).

1.2.2.4. Liquid water cooking

This method is similar to steam explosion. However, unlike steam explosion, water is maintained in liquid state at high temperature by applying pressure, which results in cooking of

the biomass. It is effective in hydrolysis of hemicelluloses and removal of lignin, accelerating the enzymatic hydrolysis of cellulose afterwards. A major advantage is that it requires less energy than steam explosion and generates less inhibitory products. However, the use of a larger volume of water results in dilution of the substrates. As a result, the downstream processing will be more energy demanding due to the need for removal of water in the later stages (Agbor *et al.,* 2011).

1.2.2.5. Dilute acid pretreatment

In dilute acid pretreatment, various acids including sulfuric acid, nitric acid, phosphoric acid and hydrochloric acid can be employed. Because of the low cost and high effectiveness, sulfuric acid is the best choice (Nguyen *et al.,* 2000). The process involves mixing of the biomass in dilute acid to solubilize the hemicelluloses and further heating at a temperature of 140 to 215°C to disrupt the crystalline structure of cellulose. In a modified two-stage acid pretreatment, a less severe condition is first used to recover the broken down hemicellulose fraction. A second more severe condition (with higher temperature and higher acid concentration) is then applied to break down the cellulose portion. By varying the pretreatment condition, significant hydrolysis of both hemicellulose and cellulose can be achieved. Advantages of dilute acid hydrolysis include effective hydrolysis of hemicelluloses, hydrolysis of a portion of the cellulose to glucose as well as improvement of further cellulose digestibility (Nguyen *et al.,* 2000). On the other hand, this method produces a substantial amount of inhibitors such as HMF and furfural depending on the severity of the pretreatment. The use of sulfuric acid also requires expensive corrosion resistant materials increasing the capital cost.

1.2.2.6. Alkali pretreatment

A typical feature of alkali pretreatment is its effective removal of the lignin fraction preserving cellulose and most of hemicelluloses in the solid part. Various methods of pretreatment employing alkali chemicals have been evaluated. The most common ones include lime pretreatment and ammonia fiber/freeze explosion (AFEX).

Lime pretreatment utilizes aqueous calcium hydroxide at low temperature and pressure for a period of a few hours to weeks. It requires fewer reagents and reduces the energy demand when performed at a temperature below 100°C. However, the long incubation time required makes it less attractive for industrial application (Chang *et al.,* 1997).

In AFEX, biomass is brought into contact with liquid ammonia for a period of 10 to 60 minutes at a temperature of about 60-100°C and pressure of above 3 MPa. An explosive relieve of pressure results in evaporation of ammonia, swelling of the biomass and decrystallization of cellulose (Mosier *et al.,* 2005). In addition, AFEX pretreatment delignifies hemicelluloses and

make it more soluble. In contrast to other pretreatment methods, AFEX allows pretreatment at very high dry matter content. Other advantages of AFEX are that only a small amount or no inhibitors at all are generated and that it requires no initial size reduction of the biomass (Zheng *et al* 2009). However, AFEX is not so effective against biomass with high lignin content such as soft wood and newspaper. Moreover, the high cost and environmental concern with the use of ammonia have a negative impact on the potential for wide scale application of this pretreatment method.

1.2.3. Enzymatic hydrolysis

Many of the pretreatment methods alter the hemicellulose fraction and release the fermentable sugar monomers. The cellulose fraction is usually not hydrolyzed to monomeric sugars but the crystalline structure is disrupted and made more accessible for enzymatic hydrolysis. For this reason, enzymatic hydrolysis of pretreated biomass usually focuses on degradation of cellulose. However, a significant fraction of hemicellulose also remains after many of the pretreatment processes, especially those using alkali pretreatment, and therefore, efficient hydrolysis of biomass should generally use a combination of both cellulose and hemicellulose degrading enzymes.

1.2.3.1. Cellulose degrading enzymes (cellulases)

Cellulases are comprised of at least three types of enzymes, endo-(1,4)-*β*-D-glucanase (EC 3.2.1.4), exo-(1,4)-*β*-D-glucanase or cellobiohydrolase (CBH) (EC 3.2.1.91), and *β*-glucosidase (EC 3.2.1.21). Endoglucanases hydrolyse *β*-1,4 linkages randomly within the cellulose chain, and create new reducing and non-reducing ends that can be acted on by CBHs. CBHs cleave cellobiose units from both reducing and non-reducing ends by moving processively along the cellulose chain. The cellobiose units together with other short chain dextrins will then be hydrolysed to monomeric glucose by the action of *β*-glucosidase (Figure 1.4). The three types of enzyme act synergistically in that, the action of endoglucanases creates new cellulose ends, which can be acted on by exoglucanases. The glucosidases also hydrolyse cellobiose, which in turn relieves possible feedback inhibition by cellobiose on CBHs (Merino & Cherry, 2007).

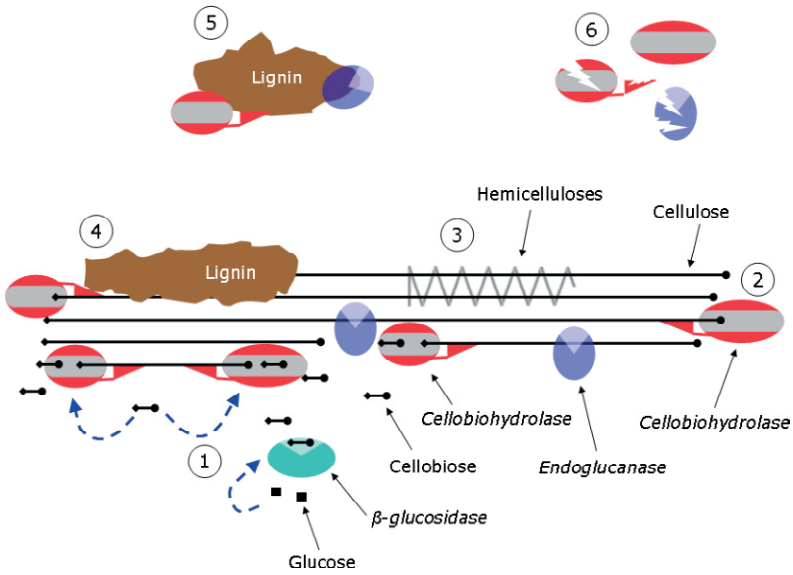


Figure 1.4. Hydrolysis and challenges in hydrolysis of cellulose (Jørgensen *et al.,* 2007). 1: glucosidase and cellobiohydrolase action and product inhibition by glucose and cellobiose, respectively. 2: binding of cellobiohydrolases onto cellulose chains. 3 and 4: hemicellulose and lignin, respectively, hinder access to cellulose for cellulase action. 5: Non-specific binding of lignin to enzymes. 6: enzyme denaturation caused by various effects of the lignocellulosic substrate.

More recently, a family of cellulolytic enzymes called polysaccharide monooxygenase (PMO) has been described. PMOs had been previously classified as glycoside hydrolase (GH61) family due to the very weak endoglucanase activity observed (Karlsson et al., 2001). However, PMOs are recently shown to cleave cellulose chain through oxidation, rather than hydrolysis (Phillips et al., 2011). This copper dependent enzyme requires not only molecular oxygen, but also reducing equivalents from cellobiose dehydrogenase (CD) that is also expressed in many fungi. PMO and CD act synergistically to augment the cellulase activity of cellobiohydrolase and endoglucanase, probably due to generation of free ends for cellobiohydrolases (Žifčáková and Baldrian, 2012). This effect of enhancing cellulase activity when used with other cellulolytic enzymes, led to the increased interest to understand the mechanism of action of PMOs, which might contribute in reducing the cost of cellulosic biofuels. However, further studies are required to unravel the role of PMOs in breaking down of lignocelluloses.

1.2.3.2. Hemicellulose degrading enzymes (Hemicellulases)

Unlike cellulases, which break down linear chains of glucose, hemicellulases are even more complex because of the need for specific enzymes that break down each kind of bond formed by the different backbones and side chains of hemicelluloses. Various hemicellulolytic enzymes have been characterized. The main chain degrading enzymes include endo-1,4-*β*-d-xylanases (EC 3.2.1.8), that break down internal bonds in the xylan chain; 1,4-*β*-D-xylosidases (EC 3.2.1.37), which chop D-xylose from the non-reducing ends of xylooligosaccharides; endo-1,4-*β*-

D-mannanases (EC 3.2.1.78), which cut internal bonds in mannan residues, and 1,4-*β*-D-mannosidases (EC 3.2.1.25), which hydrolyse mannobiose or mannooligosaccharides to mannose. Enzymes that remove side chains, called accessory enzymes are also important for efficient hemicellulose degradation. These include *α*-D-galactosidases (EC 3.2.1.22) that act on alpha-galactopyranose and mannooligomers; *α*-L-arabinofuranosidases (EC 3.2.1.55) that break alpha-arabinofuranosyl or xylooligomers; α-glucuronidases (EC 3.2.1.139), which cleave 4-O-methyl-alpha-D-glucuronic acid from 4-O-methyl-alpha-d-glucuronoxylan; acetyl xylan esterases (EC 3.1.1.72), which remove acetyl groups in xylan; and feruloyl esterases (EC 3.1.1.73), which cleave feruloyl groups from an esterified sugar polymer (Smith and Forsberg, 1991); (J. Zhang *et al.,* 2011).

1.3. Challenges in cost effective advanced bioethanol production

Complete hydrolysis of lignocellulosic biomass into fermentable sugars is a critical factor for cost effective industrial ethanol production. Though close to theoretical values can sometimes be achieved through acid hydrolysis, a number of factors, like generation of fermentation inhibitors, limit the use of this type of hydrolysis alone. Pre-treatment followed by hydrolysis with various mixtures of enzymes has been proven to be more efficient. Even then, a number of factors affect the efficiency of enzyme hydrolysis. For example, lignin that remains in the biomass after pretreatment is known to inhibit enzyme action both directly through adsorption of the enzyme and by shielding the cellulose from enzyme access. Degradation products from the pretreatment, such as organic acids, also cause inhibition of both cellulases and hemicellulases. In addition, high substrate loading can also result in product inhibition.

To achieve a higher yield of fermentable sugars from biomass, different techniques can be employed. Washing pretreated material is one possibility to remove inhibitors of both enzymatic hydrolysis and the fermentation process (Tengborg *et al.,* 2001). Another possibility is continuous removal of end products of hydrolysis that create feedback inhibition on enzymes. This is typically achieved by the use of simultaneous saccharification and fermentation (SSF). In SSF, hydrolysis of cellulose and fermentation of its product glucose are run simultaneously so that the glucose monomer formed is directly fermented to ethanol so that it does not accumulate and inhibit the glucosidase activity. Though this method has been shown to be a feasible option (Olofsson *et al.,* 2010; Prasetyo *et al.,* 2011), finding optimum conditions for both the fermenting yeast and the hydrolyzing enzymes is a challenge. This is because most of the enzymes have a temperature optimum above 50°C whereas common fermenting microorganisms have optimum temperatures below 37°C. Therefore, a compromise has to be made especially for the enzymatic hydrolysis conditions, which results in a temperature of choice close to 37°C (Olofsson *et al.,* 2008a). It was also shown that pre-fermentation of pretreated biomass spruce before SSF increases the overall yield of ethanol.

This is because pre-fermentation reduces the amount of glucose in the biomass, which in turn reduces inhibition of D-xylose uptake (Bertilsson *et al.,* 2009).

Another challenge related to cost effective industrial application of advanced bioethanol production is finding a robust organism that completely converts all the hydrolyzed sugars from cellulose and hemicellulose under industrial conditions. Some of the industrial conditions that significantly influence the economics of advanced ethanol production include process water economy, inhibitor tolerance and ethanol yield (Hahn-Hägerdal *et al.,* 2007a). Process water economy requires the use of minimum water since water has to be removed at the later stage, increasing the operational cost. Minimum water usage means high solid content and high osmolarity. Additionally, a high solid content also means a high concentration of inhibitors generated during pretreatment and hydrolysis. Therefore, both osmo- and inhibitor-tolerance are prerequisites for the organisms used in industrial application. Moreover, since biomass is a major contributor to the total production cost (Kumar & Murthy 2011), a high overall ethanol yield is crucial and this requires conversion of all the sugars present in the biomass. This necessitates strains that ferment all the sugars in the biomass, including pentose sugars (especially D-xylose, which accounts for 30-40% of total sugars) into ethanol with minimum by-product formation.

In practice, there is no organism that meets all these requirements. Each group of organisms has its own advantages and disadvantages with respect to certain characteristics. For example, the yeast *S. cerevisiae* ferments hexose sugars close to theoretical maximum yield and has significant tolerance to various inhibitors including ethanol. However, it is not able to naturally ferment pentose sugars. Many species of bacteria, yeasts and filamentous fungi are able to ferment the pentose sugars D-xylose and arabinose into ethanol. However, these organisms suffer from a number of drawbacks including low tolerance to ethanol and inhibitors from lignocellulosic hydrolysates.

1.4. Pentose metabolism in fungi and bacteria

Bacteria and fungi employ different D-xylose metabolism pathways. In most naturally D-xylose utilizing fungi, D-xylose is first reduced to xylitol by NAD(P)H dependent Xylose Reductase (XR) encoded by the gene *XYL1*. Xylitol is then oxidized to xylulose by NAD^+ dependent Xylitol Dehydrogenase (XDH) encoded by *XYL2*. Xylulokinase (XK) subsequently phosphorylates Xylulose into Xylulose-5-phosphate, an intermediate of the pentose phosphate pathway (PPP), which connects it to the central metabolism. In bacteria, D-xylose is directly isomerized to Xylulose by the *xylA* encoded enzyme, Xylose Isomerase (XI). Similar to the fungal pathway, xylulose is phosphorylated to xylulose-5-phosphate, which is channeled into the central metabolism through the pentose phosphate pathway (Figure1. 5). The D-xylose metabolism

pathway involving XI has also been reported in some filamentous fungi and yeast (Banerjee *et al.,* 1994; Harhangi *et al.,* 2003).

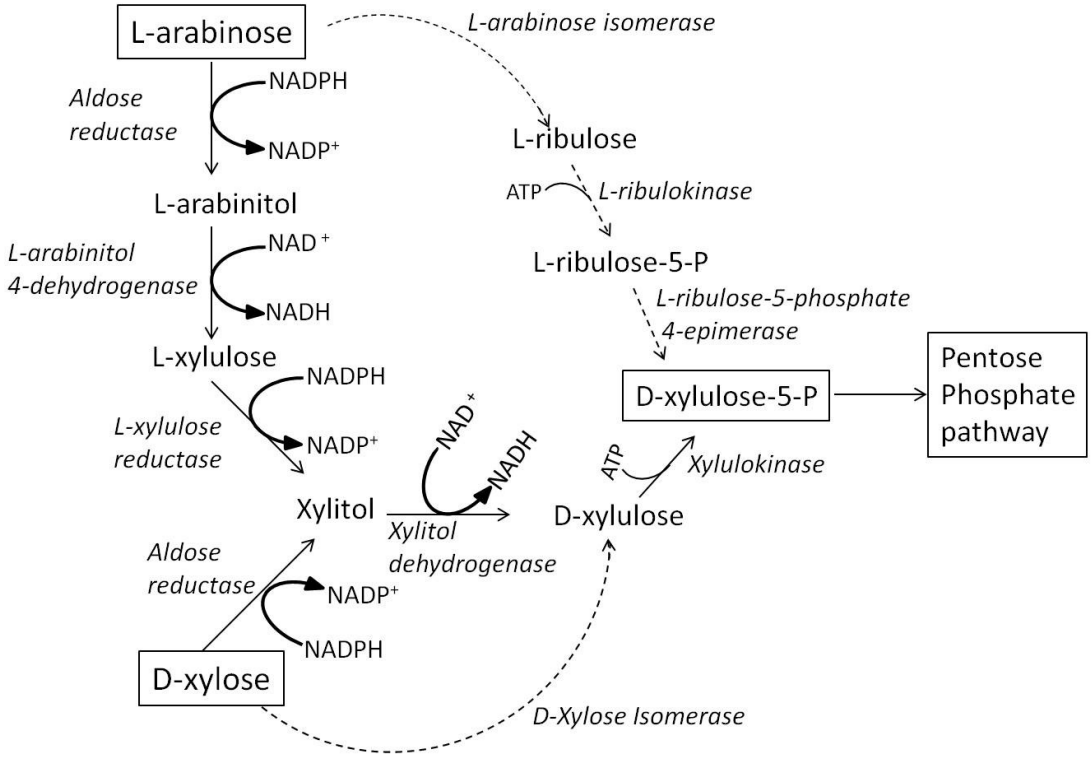


Figure 1.5. Pentose metabolism in fungi and bacteria. The oxidoreductase pathways are shown in solid arrows while isomerization pathways are indicated in broken line arrows. In yeast and filamentous fungi, D-xylose is converted to xylulose-5-phosphate by D-xylose/aldose reductase (*XYL1*), xylitol dehydrogenase (*XYL2*), D-xylulokinase (*XYL3*). In bacteria and some fungi, xylose isomerase (*xylA*) bypasses the oxidation/reduction steps. Arabinose in fungi, is converted to D-xylulose in 4 steps using enzymes xylose/aldose reductase (*XYL1),* L-arabinitol 4-dehydrogenase (*lad1),* L-xylulose reductase (lxr1), xylitol dehydrogenase (*XYL2*). In bacteria arabinose is metabolized by three enzymes; L-arabinose isomerase (araA), L-ribulokinase (araB), and L-ribulose-5-phosphate 4-epimerase (araD) to Xylulose-5-Phosphate. Xylulose-5-phosphate enters into the pentose phosphate pathway.

Similar to D-xylose metabolism, bacteria and fungi have distinct L-arabinose metabolism pathways (Figure 1.5). In fungi, L-arabinose is reduced to L-arabinitol by the NADPH dependent xylose/aldose reductase encoded by *XYL1.* L-arabinitol is subsequently oxidized to L-xylulose by the NAD dependent L-arabinitol 4-dehydrogenase, encoded by *lad1*. In the second redox reaction, L-xylulose is reduced to xylitol by L-xylulose reductase (encoded by lxr1), and xylitol is oxidized to D-xylulose by xylitol dehydrogenase (ecoded by *XYL2*). D-xylulose is then phosphorylated by xylulokinase to D-xylulose-5-phosphate, an intermediate of the pentose phosphate pathway. In bacteria, L-arabinose is converted to D-xylulose-5-phosphate in three steps enzymatic reaction that involves: 1) isomerization of L-arabinose to L-ribulose by L-arabinose isomerase (encoded by *araA*), 2) phosphorylation of L-ribulose to L-ribulose-5-

phosphate by L-ribulokinase (encoded by *araB*), and 3) epimerization of L-ribulose-5-phosphate to D-xylulose-5-phosphate by L-ribulose-5-phosphate 4-epimerase (encoded by *araD*). Since the fungal oxidoreductase L-arabinose pathway generate redox imbalance, most L-arabinose utilizing fungi do not ferment L-arabinose into ethanol, but accumulate intermediate products such as arabinitol. As a result most strain engineering strategies focused on expression of the bacterial L-arabinose pathway into the yeast *S. cerevisiae* (McMillan and Boynton, 1994; Wisselink et al., 2007).

Since D-xylose is the most abundant pentose in lignocellulosic hydrolysates (Table 1), the focus of this thesis is on the construction of D-xylose utilizing strain. Therefore, a more detailed description of the D-xylose metabolism and the strategies to improve D-xylose fermentation is made in the following sections.

1.4.1. Natural D-xylose utilizing microorganisms

Several species of yeast that fermentatively metabolize D-xylose exist. However, only about 1% of them convert D-xylose to ethanol. The yeasts *Pichia stipitis, Candida shehatae, and Pachysolen tannophilus* are among the well known organisms than naturally convert D-xylose into ethanol (Jeffries and Jin, 2000). *P. stipitis* has acquired wide spread acceptance as a promising candidate organism owing to its high rate of D-xylose fermentation (Agbogbo and Coward-Kelly, 2008).

1.4.1.1. *Pichia stipitis (Scheffersomyces stipitis)*

P. stipitis is one of the promising candidate organisms for advanced bioethanol production. It is closely related to yeasts that are isolated from the gut of passalid beetles, which inhabit and degrade white-rotted hardwood (Suh *et al.,* 2003). This ecological niche is associated with the ability of the organism to utilize most of the sugars present in woody biomass. *P. stipitis* ferments not only all the common hexose and pentose sugar monomers in lignocellulosic biomass, but also the disaccharide cellobiose (Agbogbo and Coward-Kelly, 2008). Cellobiose is an important feedback inhibitor of cellulose hydrolysis. This yeast can therefore be used especially in SSF without the need to use extra beta-glucosidase.

One of the biggest drawbacks of this species is the lack of growth in anaerobic conditions even with hexose sugars. Moreover, unlike in *S. cerevisiae*, a high sugar concentration does not induce fermentation in *P. stipitis.* The optimum condition for ethanolic fermentation in this species is the micro-aerobic condition. However, controlled oxygenation is practically not feasible in large-scale industrial conditions (Hahn-Hägerdal *et al.,* 2007). In addition, the rate of sugar fermentation in *P. stipitis* is generally much lower than that of *S. cerevisiae*. Furthermore, this species has lower tolerance to ethanol and to the inhibitors present in lignocellulosic hydrolysates. Currently, good progress has been made to improve the sugar fermentation rate,

tolerance to various inhibitors and to understand the metabolism under different oxygen levels (Skoog and Hahn-Hägerdal, 1990; Unrean and Nguyen, 2012; Yong *et al.,* 2012). However, there is still much to be done before this yeast can be used industrially.

1.4.1.2. Filamentous fungi

Filamentous fungi have been used in industry for production of various metabolites and enzymes. Various species of filamentous fungi such as *Fusarium, Aspergillus,* and *Trichoderma*, are known to degrade both cellulose and hemicellulose and convert both hexose and pentose sugars into ethanol (Jun *et al.,* 2011). Therefore, such species are considered as very good candidates for consolidated bioprocessing (CPB), one step conversion of lignocelluloses into the desired product without external addition of cellulases and hemicellulases. In particular, *T. reesei* is superior to other filamentous fungi in production of a high titer of ethanol from cellulose. On the other hand, ethanol production from monomeric sugars is lower in *T. reesei* compared to other fungi. Even though most of the strain development strategies for CBP focus on modification of *S. cerevisiae* for secretion of cellulases and hemicellulases, efforts to engineer natural lignocellulose degrading fungi have also been evaluated (Xu *et al.,* 2009). The secretion of large amounts of cellulases by *T. reesei,* its ability to utilize all sugars in the biomass and the proven record for industrial application, make it a very good candidate for a future CBP strain. However, several challenges have to be addressed before this organism can be used as a CBP strain. These challenges include that *T. reesei* has a lower tolerance to ethanol and lower ethanol yield and productivity compared to *S. cerevisiae*. In addition, *T. reesei* is an obligate aerobe and has a filamentous morphology. Hence, it requires continuous oxygen supply and mixing within the fermentation tank, which requires a substantial amount of energy (Xu *et al.,* 2009).

1.4.1.3. Bacteria

Bacteria, such as *Escherichia coli, Klebsiella oxytoca*, and some thermophilic bacteria such as *Clostridium phytofermentans,* readily ferment all sugars present in biomass. However, these organisms predominantly produce various acids, and ethanol is only a minor product. Though successful efforts have been undertaken to divert the metabolism towards ethanol production, the tolerance of these organisms to ethanol and other inhibitors is far below industrial requirements (Shaw *et al.,* 2008). In addition, the optimum pH for bacteria is close to neutral, which allows easy contamination of the fermentation process.

E.coli strains that principally produce ethanol as a fermentation product were one of the earliest successes of metabolic engineering (Ingram *et al.,* 1987). Later on, several strains which produce ethanol as sole fermentation product have been developed. A significant challenge remains the development of strains with higher tolerance to ethanol and to lignocellulosic

inhibitors. In this respect, new strains continue to be reported, but the performance of the ethanologenic *E. coli* remains inferior to that of *S. cerevisiae* with respect to yield and productivity (Lau *et al.,* 2010).

1.4.2. Other ethanologenic microorganisms

1.4.2.1. *Zymomonas mobilis*

Z. mobilis is among the potential candidates for industrial ethanol production owing to its high rate and yield of ethanol production from glucose with no by-product formation. It also displays good tolerance to ethanol. However, *Z. mobilis* has a very limited substrate utilization range (glucose, fructose and sucrose) and low tolerance towards weak acids and other inhibitors generated during pretreatment of lignocellulosic hydrolysates. Several recombinant strains that circumvent these problems have been constructed through metabolic engineering. The construction of recombinant strains that convert D-xylose into ethanol was successful through expression of bacterial xylose isomerase and xylulokinase, together with the pentose phosphate pathway gene *tktA,* coding for transketolase (Agrawal *et al.,* 2011; Zhang *et al.,* 1995). In addition, the tolerance of a recombinant D-xylose fermenting strain for acetate has been significantly improved (Agrawal *et al.,* 2012, 2011). However, the rate of D-xylose fermentation is significantly lower compared to that of glucose. Furthermore, much work still has to be done to obtain a robust strain that can withstand the various inhibitors in lignocellulosic hydrolysates before it is possible to use this organism in industrial application.

1.4.2.2. *Saccharomyces cerevisiae*

S. cerevisiae has been traditionally used in large scale industrial applications including production of alcoholic beverages, as baker's yeast, and for bioethanol production. In the current sugar and starch based bioethanol industries, *S. cerevisiae* is used as a sole production organism. This is because this organism produces ethanol from hexose sugars with stoichiometric yield, and demonstrates substantial tolerance to a wide range of inhibitors. Furthermore, *S. cerevisiae* is one of the most intensively studied microorganisms, and numerous tools exist for genetic manipulation of the organism. The major drawback for using *S. cerevisiae* for advanced bioethanol production is its inability to ferment the pentose sugars D-xylose and L-arabinose to ethanol (Hahn-Hägerdal *et al.,* 2007a). Because of the possibility of integrating it into the already existing starch and sugar based ethanol industries, engineering *S. cerevisiae* for advanced bioethanol production has gained widespread attraction.

1.5. Engineering *S. cerevisiae* for D-xylose fermentation

As described above, *S. cerevisiae* cannot metabolize D-xylose. Although genes for D-xylose utilization exist in *S. cerevisiae*, the expression of these genes is too low to support fermentation or even growth in aerobic conditions. Because of the inherent advantage of using

S. cerevisiae as a robust industrial strain, substantial efforts and much progress have been made to engineer this yeast for D-xylose utilization, compared to engineering natural D-xylose utilizing strains for various industrial properties that *S. cerevisiae* has already acquired. Engineering D-xylose utilization in *S. cerevisiae* has been achieved by expressing either the XI or the XR/XDH pathways from other D-xylose assimilating organisms (Chu and Lee, 2007). In addition to such a rational metabolic engineering approach, alternative inverse metabolic engineering and evolutionary engineering strategies have also been proven successful.

1.5.1. Rational metabolic engineering

Metabolic engineering can be defined as the intentional redirection of a certain metabolic flux to a desired product formation or to new cellular properties through modification or introduction of specific biochemical pathways using recombinant DNA technology (Nevoigt, 2008). It has been intensively applied for the production of several products in the food and beverage industry, biofuel, and fine and bulk chemical production. In contrast to other strain improvement strategies like evolutionary engineering and classical mutagenesis, metabolic engineering has two major advantages. First, the desired trait can be specifically modified without introduction of unnecessary mutations. Second, new phenotypes can be generated through the introduction of foreign pathways that were not present in the target strain (Nevoigt, 2008). The most important targets of metabolic engineering for D-xylose fermentation in *S. cerevisiae* focused on D-xylose uptake and on the initial assimilation steps.

1.5.1.1. D-xylose uptake

S. cerevisiae takes up D-xylose through its native hexose transporters (HXT family), which have 10-100 times less affinity for D-xylose compared to that of glucose (Saloheimo *et al.,* 2007). In addition, the high affinity hexose transporters that also transport D-xylose are repressed in the presence of glucose. This affects the rate of D-xylose utilization in yeast strains engineered for D-xylose metabolism, especially strains with a higher level of expression of the initial D-xylose assimilation genes. Thus, simultaneous utilization of D-xylose with hexose sugar is hampered mainly due to competitive inhibition of D-xylose uptake by glucose (Subtil and Boles, 2012).

To circumvent these problems, several efforts have been made through over-expression of endogenous sugar transporters and expression of heterologous sugar transporters in *S. cerevisiae*. The endogenous hexose transporters *HXT1/2/4/5/7* and *GAL2* have been shown to effectively transport D-xylose in a D-xylose utilizing *S. cerevisiae* strain. While *HXT1* and *HXT4* are expressed in the presence of glucose, *GAL2* is expressed in the presence of galactose but in the absence of glucose. However, *HXT5* and *HXT7* are constitutively expressed independent of the presence of glucose, showing that these two transporters are good targets for strain engineering in order to enhance D-xylose transport (Hamacher *et al.,* 2002; Sedlak & Ho 2004).

Previous studies have shown that over-expression of the native D-xylose transporters (like those from the HXT family) in D-xylose utilizing strains results in no or only little improvement of the rate of D-xylose utilization (Hamacher *et al.,* 2002; Sedlak and Ho, 2004b). However, these experiments were performed in strains displaying only slow D-xylose utilization capacity. Recently, over-expression of *HXT7* and *GAL2* has been shown to enhance the D-xylose fermentation rate in *S. cerevisiae* strains with expression of bacterial XI that display much faster D-xylose utilization (Subtil and Boles, 2012). Nevertheless, D-xylose was consumed only after glucose was exhausted from the medium. Therefore, expression of heterologous D-xylose transporters that specifically transport D-xylose maybe crucial for faster D-xylose fermentation, especially for co-consumption with glucose.

A number of heterologous D-xylose transporters with high affinity for D-xylose have been expressed in *S. cerevisiae*. The *P. stipitis SUT1*, the *Candida intermedia GXS1* and *GXF1*, the *Arabidopsis thaliana At5g59250* transporters were shown to efficiently transport D-xylose when expressed in yeast. However, the heterologous transporters expressed so far exhibited lower maximal uptake activity compared to the over-expressed native *HXT* transporters. Moreover, the glucose/xylose-H^+ symporter *GXS1* is an ATP requiring transporter. Hence, it may not be very useful since only a limited amount of ATP is produced during anaerobic fermentation, although this may also result in a higher fermentation rate to meet the higher demand for ATP. To date no efficient and specific D-xylose transporter that cannot be repressed by glucose has been successfully expressed in *S. cerevisiae* (Hector *et al.,* 2008; Leandro *et al.,* 2006).

1.5.1.2. Conversion of D-xylose to xylulose

1.5.1.2.1. The XR/XDH pathway

The yeast *S. cerevisiae* is able to metabolize D-xylulose to ethanol (Figure 1.6). Though the genes encoding enzymes able to convert D-xylose to D-xylulose are present in the genome, they do not support growth on D-xylose as sole carbon source. Over-expression of the endogenous genes, *GRE3,* coding for non-specific aldose reductase that converts D-xylose to xylitol, and *XYL2,* coding for xylitol dehydrogenase (XDH) that converts xylitol to D-xylulose has been shown to allow growth in D-xylose medium. However, the rate of growth and fermentation was very low compared to strains with heterologous expression of *XYL1,* coding for xylose Reductase (XR), and *XYL2* from other yeast species (Toivari *et al.,* 2004).

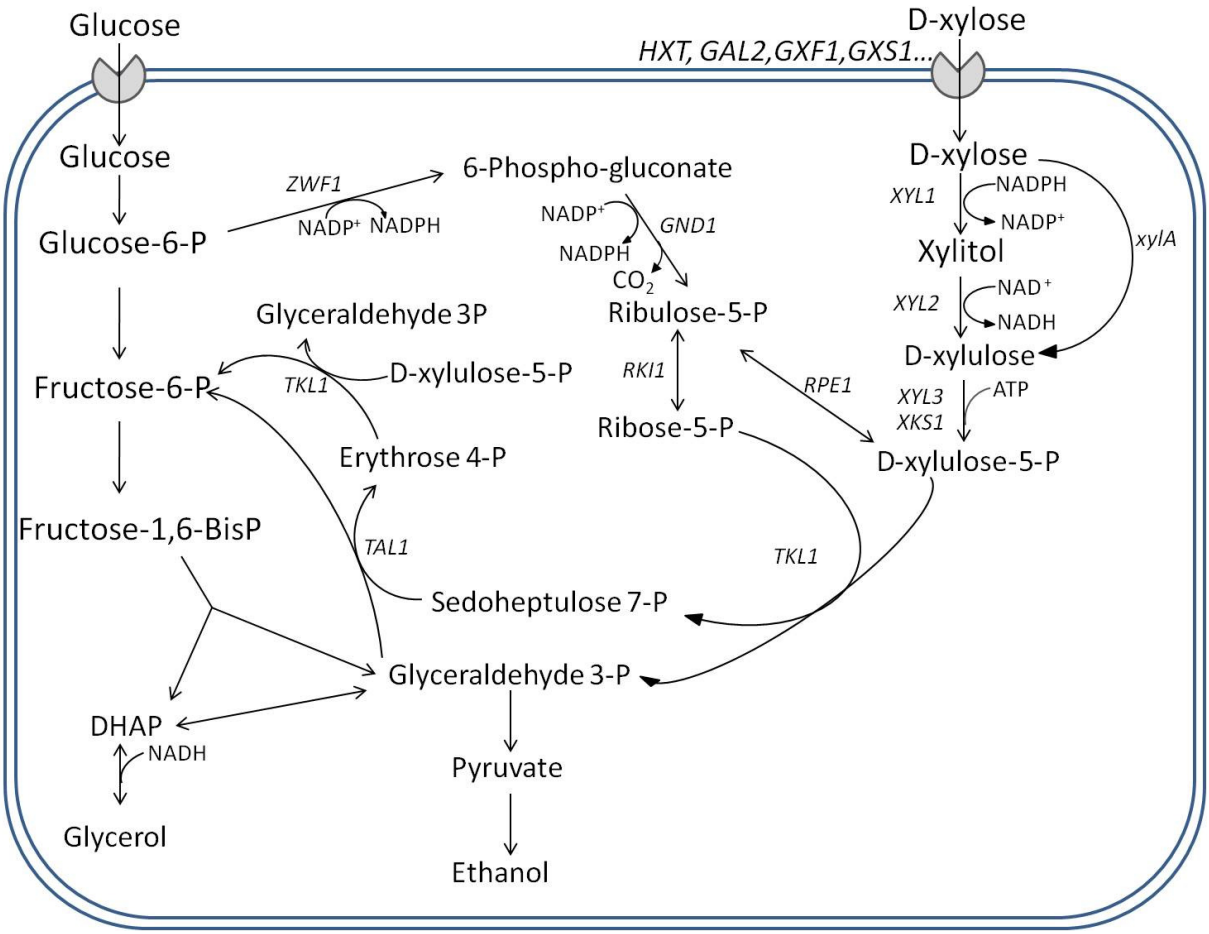


Figure 1.6. Overview of the engineering strategies for D-xylose metabolism in *S. cerevisiae*. To enhance D-xylose uptake, endogenous hexose transporters (HXT, *GAL2*) or heterologous D-xylose transporters *GXF1* or *GXS1* are overexpressed. Conversion of D-xylose to D-xylulose is achieved by heterologous expression of D-xylose reductase (*XYL1*), xylitol dehydrogenase (*XYL2*) from natural D-xylose utilizing yeast; or xylose isomerase (*xylA*) from bacteria or some fungi. Further improvement of the metabolism rate downstream of D-xylulose is achieved by over-expression of D-xylulokinase (*XYL3* or *XKS1*(endogenous)) and the PPP enzymes D-ribulose-5-phosphate 3-epimerase (*RPE1*), ribose-5-phosphate ketol-isomerase (*RKI1*), transketolase (*TKL1*) and transaldolase (*TAL1*).

In fact, the first *S. cerevisiae* strain able to utilize D-xylose anaerobically was constructed by expression of XR and XDH from *P. stipitis* (Eliasson *et al.,* 2000). Later, several such strains have been developed (Jin and Jeffries, 2003; Toivari *et al.,* 2001). However the efficiency of D-xylose fermentation in those recombinant strains was considerably lower compared to that of *P. stipitis*, primarily due to differences in cofactor preference of XR and XDH, resulting in accumulation of xylitol instead of ethanol. The *P. stipitis* XR prefers NADPH over NADH. On the other hand, XDH uses only NAD^+ and not $NADP^+$ as a cofactor. This leads to accumulation of NADH since the yeast cell cannot recycle excess NADH during anaerobic fermentation. As a result, a fraction of D-xylose will be converted to and accumulated as xylitol and the ethanol yield is reduced (Eliasson *et al.,* 2000; Jeffries, 2006; Karhumaa *et al.,* 2005) .

A number of studies have been conducted to solve the problem of cofactor imbalance through modification of the cofactor specificity of XR and XDH. Several XRs with high coenzyme preference for NADH rather than NADPH have been generated. However, in most cases, the increased preference for NADH was the result of a decrease in the catalytic efficiency with NADPH and therefore the result was an overall reduction in ethanol productivity (Liang *et al.,* 2007; Watanabe *et al.,* 2007). More recently, a recombinant strain expressing a mutated XR with altered coenzyme preference for NADH showed a significant increase in ethanol production from D-xylose in both mineral medium and acid pretreated wheat straw hydrolysate (Runquist *et al.,* 2010; Olofsson *et al.,* 2011). Yet, the ethanol yield from D-xylose was very low compared to the theoretically expected yield.

Alteration of coenzyme specificity of XDH from NAD^+ to $NADP^+$ has also shown promising results. Watanaba *et al.,* (2005) have successfully changed the cofactor specificity of XDH to $NADP^+$ through site-directed mutagenesis. The resulting strain showed significantly improved ethanol yield and productivity while the xylitol yield was very low compared to the strain expressing the wild type gene for XDH (Matsushika *et al.,* 2009; Watanabe *et al.,* 2005).

The problem of cofactor imbalance in XR/XDH expressing strains of *S. cerevisiae* has also been addressed through heterologous expression of bacterial transhydrogenases. Transhydrogenases catalyze the transfer of H^+ between the two pyridine nucleotide systems, $NADH/NAD^+$ and $NADPH/NADP^+$, and occur in various organisms but not in *S. cerevisiae*. When the bacterial transhydrogenase from *Azotobacter vinelandii* was expressed in *S. cerevisiae* harboring the XR/XDH system, the yield of xylitol was indeed reduced, but the strain produced more glycerol instead of more ethanol (Jeppsson *et al.,* 2003; Nissen *et al.,* 2001).

Overall, balancing the redox potential in yeast expressing the XR/XDH system has shown to improve ethanol yield in some cases but could not eliminate xylitol production in order to maximize ethanol yield. In some cases, sugar metabolism did not proceed in the direction of ethanol but rather to other products like glycerol or acetate (Jeppsson *et al.,* 2003). In other instances, the strains with balanced redox potential still did not show a significant decrease in xylitol yield indicating that other unknown metabolic pathways might be responsible for accumulation of xylitol with lower ethanol yield as a result.

Formation of the by-product xylitol is also dependent on the strain background and the fermentation medium used. For example, some strains of *P. stipitis* were shown to produce no xylitol at all, irrespective of the level of oxygenation (Skoog and Hahn-Hägerdal, 1990). With respect to the fermentation medium, it has been shown that fermentation of lignocellulosic hydrolysate by XR/XDH carrying strains resulted in significantly lower or even no xylitol

accumulation at all, compared to fermentations in synthetic medium (Karhumaa *et al.,* 2007b; Sedlak and Ho, 2004a). The absence of xylitol during fermentation of undetoxified lignocellulosic hydrolysate can be explained by the presence of high levels of external electron acceptors in this medium, such as furfural, which is known to support oxidation of NADH to NAD^+.

1.5.1.2.2. The Xylose Isomerase (XI) pathway

Recombinant strains of *S. cerevisiae* engineered with the XR/XDH pathway mostly show low ethanol yield and accumulate xylitol, which is likely due to the redox imbalance. The fact that *S. cerevisiae* metabolizes xylulose, and the existence of other organisms, like bacteria, that directly isomerize xylose to xylulose by the enzyme XI, led some researchers to express this enzyme into yeast. Addition of XI into medium containing D-xylose resulted in a slow fermentation of D-xylose by yeast that does not express XR (Bruinenberg *et al.,* 1983). However, expression of bacterial XI into yeast was not as easy as anticipated, and for many years, no functionally active XI could be expressed in yeast (Gárdonyi and Hahn-Hägerdal, 2003).

The first functional expression of prokaryotic XI in yeast was achieved using the x*ylA* gene from the thermophilic bacterium *Thermus thermophilus* (Walfridsson *et al.,* 1996). Though the highest activity of this enzyme was at 85°C, the recombinant strain was able to slowly grow and ferment D-xylose to ethanol at 30°C. Further adaptation of *xylA* through random mutagenesis resulted in thermal adaptation of the XI to lower temperature (Lönn *et al.,* 2002). Moreover, additional genetic modifications of the strain expressing *T. thermophilus* XI resulted in improved growth and fermentation of D-xylose. However, the rate of D-xylose fermentation by the adapted strain was very low compared to strains constructed by expression of the XR/XDH pathway (Karhumaa *et al.,* 2005).

After the discovery of the first fungal XI in a strain of *Piromyces* sp. strain E2 (Harhangi *et al.,* 2003), the enzyme was functionally expressed in *S. cerevisiae* and showed high enzymatic activity. The success in functional expression of the fungal XI might be due to the close relationship of this eukaryote with *S. cerevisiae*, which has similar cellular processes like protein folding and cytosolic environment. On the other hand, the high activity of the enzyme in *S. cerevisiae* supported only slow growth and fermentation in D-xylose medium, suggesting the possible involvement of other metabolic pathways necessary for faster D-xylose growth and fermentation (Kuyper *et al.,* 2003a).

Since then, other bacterial XIs have also been expressed in *S. cerevisiae*. The *Bacteroides thetaiotaomicron* and *Xanthomonas campestris* XI, which showed high similarity to the

Piromyces XI, were actively expressed in *S. cerevisiae*. However, the activities of these bacterial enzymes in *S. cerevisiae* were lower compared to that of *Piromyces* XI (Hahn-Hägerdal *et al.,* 2007b). More recently, the first bacterial XI showing high activity in *S cerevisiae* was successfully expressed. This recombinant enzyme originated from the anaerobic bacterium *Clostridium phytofermentans* and was less inhibited by xylitol compared to that of *Piromyces* XI. Further modification of the gene for the same codon usage as that of the highly expressed glycolytic genes in *S. cerevisiae* resulted in a significant improvement in the expression of the gene. Nevertheless, recombinant strains harboring the codon optimized gene from *C. phytofermentans* still showed slow growth and poor fermentation with D-xylose even after further adaptation (Brat *et al.,* 2009).

The activity of XI obtained in the recombinant *S. cerevisiae* strains constructed so far provided a promising starting point for further strain improvement. Theoretically, the XI approach seems to be superior to the XR/XDH approach since it involves direct isomerization of D-xylose to xylulose without cofactor requirement. Indeed, the XI expressing strains displayed a higher yield of ethanol per consumed D-xylose compared to strains harboring the XR/XDH pathway. On the other hand, ethanol productivity by XI expressing strains is inferior. This low productivity has been attributed partly to insufficient activity of the XI enzyme in *S. cerevisiae* (Bettiga *et al.,* 2008). Furthermore, the XI activity was dependent on the strain background in which the gene was expressed (Brat *et al.,* 2009). As a consequence, the conversion of D-xylose to xylulose alone could not support efficient D-xylose metabolism with both pathways up to now. Extensive modifications to increase the metabolic flux, and/or further evolutionary adaptation were required to improve the D-xylose utilization efficiency.

1.5.1.3. Modulating flux downstream of xylulose

Once the conversion of D-xylose to xylulose was possible in *S. cerevisiae* through heterologous expression of either XI or XR/XDH enzymes, limitations of D-xylose fermentation were thought to reside downstream of xylulose, either in the reaction catalyzed by xylulokinase (XK) or in the Pentose Phosphate Pathway (PPP). Because the glycolytic pathway is already highly active in *S. cerevisiae*, it is unlikely that this pathway limits the fermentation rate (van Maris *et al.,* 2007). Therefore, efforts to modulate the flux focused on up-regulation of xylulokinase encoded by *XKS1* and enzymes of the PPP (Figure1.6).

1.5.1.3.1. Xylulokinase (XK)

S. cerevisiae naturally harbors the gene *XKS1* (*XYL3*) that codes for the enzyme XK. However, endogenous XK activity is not enough to support rapid D-xylose utilization in recombinant *S. cerevisiae.* This was experimentally confirmed by fine-tuning the expression of XR and XDH to have a higher expression of XDH in recombinant strains expressing the *P. stiptis* XR/XDH. The

strain produced a lower amount of xylitol and accumulated a significant amount of xylulose during D-xylose fermentation, indicating that the XK activity was not high enough to support the conversion of xylulose to xylulose-5-phosphate at a rate suitable for rapid D-xylose fermentation (Jin and Jeffries, 2003). Overexpression of XK proved to be successful in rendering faster xylulose clearance by recombinant strains expressing XR/XDH from *P. stipitis* (Ho *et al.,* 1998; Toivari *et al.,* 2001). However, fine-tuned expression of *XKS1* was necessary since uncontrolled expression resulted in poor growth on D-xylose. This poor growth or even cell death associated with overexpression of XK could be explained by accumulation of xylulose-5-phosphate (X-5-P) and depletion of ATP. The PPP in *S. cerevisiae* might not be active enough to support a high flux of X-5-P and compensate for the rapid ATP consumption by XK (Jin *et al.,* 2003; Johansson *et al.,* 2001). In addition, metabolic control analysis has shown that in many cases modulation of a pathway through up-regulation of a single enzyme has little success. Therefore, simultaneous over-expression of all enzymes of D-xylose metabolism was necessary. As a result, metabolic engineering for D-xylose fermentation nowadays includes overexpression of all five genes downstream of xylulose: *XKS1* and the four enzymes of the non-oxidative PPP (Van Maris *et al*, 2007).

1.5.1.3.2. The Non-oxidative Pentose Phosphate Pathway (noPPP)

There are four main enzymes in the noPPP; ribulose-5-phosphate epimerase (*RPE1*), ribose-5-phosphate ketol isomerase (*RKI1*), transaldolase (*TAL1*) and transketolase (*TKL1*) (Figure 1.6). The metabolic flux through the noPPP in *S. cerevisiae* is considerably lower than that of *P. stipitis*. This low flux is believed to be the result of long term adaptation of *S. cerevisiae* to ethanolic fermentation of hexose sugars (Hahn-Hägerdal *et al.,* 2007b).

Since xylulose can only be channeled into the central metabolism through this pathway, it is necessary to overexpress the enzymes of the noPPP. The two main enzymes *TKL1* and *TAL1* have initially been implicated as the limiting steps for D-xylose fermentation. This implication was supported by experimental evidence that showed elevated levels of *TKL1* and *TAL1* genes in gene expression analysis of *S. cerevisiae* fermenting D-xylose (Toivari *et al.,* 2004). On the other hand, it was shown that overexpression of the four enzymes of the noPPP in a strain expressing *XYL1* and *XYL2* and overexpressing *XKS1*, doubles the rate of xylulose consumption compared to a strain expressing only *XYL1* and *XYL2* and overexpressing *XKS1* (Fiaux *et al.,* 2003). Similarly, overexpressing *XKS1* and the four noPPP genes in a recombinant strain overexpressing the *Piromyces* sp *xylA*, resulted in significant improvement of anaerobic fermentation of D-xylose to ethanol. This strain did not accumulate xylulose during growth on D-xylose (Kuyper *et al.,* 2005b). Other studies also confirmed the importance of overexpression of the noPPP enzymes together with *XKS1* in recombinant strains expression either the XI or XR/XDH pathway (Brat *et al.,* 2009; Karhumaa *et al.,* 2007a, 2005).

1.5.1.4. Other modifications

In general, the main rate limiting steps in D-xylose fermentation in recombinant *S. cerevisiae* have been attributed to insufficient D-xylose uptake, insufficient acitivty of XI in XI expressing strains or redox imbalance associated with the heterologous expression of XR/XDH enzymes, and insufficient activity of XK and PPP enzymes. In spite of extensive research to resolve these factors, no strain that ferments D-xylose to ethanol at the theoretically expected yield has been generated yet. Many other factors have been addressed to increase the ethanol yield and productivity and to reduce by-product formation, especially xylitol accumulation.

In recombinant strains expressing XI, deletion of *GRE3* that codes for the enzyme aldose reductase was proven successful. Aldose reductase catalyses the reduction of D-xylose to xylitol. Xylitol is accumulated at the expense of ethanol yield, and in addition, it inhibits the activity of XI. Therefore *GRE3* deletion reduces the formation of xylitol and increases the ethanol yield (Kuyper *et al.,* 2005a). Unfortunately, aldose reductase is one of the stress induced enzymes that are responsible for tolerance to several yeast inhibitors (Aguilera and Prieto, 2001). Therefore, deletion of *GRE3* is probably not an option in a yeast strain for industrial application. The expression of the XI from *C. phytofermentans* that is less inhibited by xylitol is therefore a very promising starting point for establishment of this pathway in industrial strains (Brat *et al.,* 2009).

Another modification is the disruption of the electron transfer chain to generate a respiratory deficient strain. The reason for this is that oxygen consumption during fermentation might result in respiration dependent ethanol utilization, reducing the overall ethanol yield. In addition, re-oxidation of NADH through respiration is necessary for higher ethanol productivity and reduced xylitol yield in XR/XDH expressing strains. This controlled oxygen supply is not economically feasible. To overcome this, deletion of genes that encode a cytochrome c oxidase subunit, like *COX4,* has been shown to eliminate respiratory growth on D-xylose in both XR/XDH and XI expressing strains. However, the respiratory deficient strains were not able to grow in D-xylose medium. Only after extensive evolutionary adaptation, the respiratory deficient strains exhibited improved growth and fermentation of D-xylose and gave a higher ethanol yield. Yet the yield of ethanol from D-xylose was still far from the maximum theoretical ethanol yield (Peng *et al.,* 2012; Shen *et al.,* 2012).

In summary, rational metabolic engineering aimed at direct modification of a certain metabolic flux was proven successful in developing strains with a new phenotype (such as D-xylose fermentation), as discussed above. A major challenge in rational metabolic engineering is the complexity of the biochemical pathways that govern cellular functions. Extensive knowledge of the metabolic pathway is required, and even then, secondary effects of a specific genetic

modification are hardly predictable. These secondary effects may occur in pathways that are unrelated to the pathway under study in which the modifications were introduced (Sauer 2001). One of the many examples is the over-expression of XK in recombinant D-xylose utilizing strains (Section 1.5.1.3.1). This was initially thought to enhance the flux of D-xylose assimilation through faster phosphorylation of xylulose to X-5-P, which is channeled into the PPP. However, the over-expression of XK was found to inhibit cell growth in D-xylose medium and only moderate activity of XK was found to be required to support growth on D-xylose (Jin *et al.,* 2003). Furthermore, in spite of good knowledge of the targeted metabolic pathway in a certain strain or species, genetic engineering in another strain or species using this information often does not give the expected results. For this reason, many of the metabolic engineering strategies produce satisfactory results only after further evolutionary engineering.

1.5.2. Evolutionary engineering

Evolutionary engineering is a method of random strain improvement using either one or a combination of the methods of random mutagenesis, genome shuffling and adaptive evolution under selective pressure. Combined with the traditional metabolic engineering, evolutionary engineering has shown great success in improving strains for the desired property (Table 2). Sonderegger and Sauer, (2003) successfully developed the first *S. cerevisiae* strain TMB3001C5 capable of growing anaerobically in D-xylose through systematic evolutionary adaptation. The original recombinant strain TMB3001 that expressed the *P. stipitis* D-xylose utilization pathway (XR/XDH) was unable to grow anaerobically on D-xylose. After extensive adaptation, first for efficient aerobic growth, then for growth under semianaerobic conditions and finally for growth under anaerobic conditions in D-xylose medium, a stable anaerobic D-xylose growing strain could be developed. Similarly, when the fungal XI was functionally expressed in *S. cerevisiae,* only moderate D-xylose utilization was accomplished by the recombinant strain even after over-expression of *XKS1* and all the genes of the non-oxidative PPP. However, evolutionary adaptation of the engineered strain in sequential batch cultures followed by growth in a D-xylose-limited chemostat culture resulted in significant improvement in both the rate of growth and yield of ethanol from D-xylose (Zhou *et al.,* 2012).

When Brat *et al.,* (2009) constructed the XI expressing strains in both laboratory and industrial strain backgrounds, only the laboratory strain was able to slowly grow in D-xylose even after the XI encoding gene was optimized for codon usage. After serial transfer in D-xylose medium, the recombinant industrial strain was able to grow slowly in D-xylose medium. This result also indicated that the success of metabolic engineering depends on the strain background used.

In another example, a metabolically engineered strain TMB3399, that was constructed by chromosomal integration of the *P. stipitis* XR and XDH together with over-expression of XK,

showed only moderate growth and fermentation of D-xylose. After chemical mutagenesis using ethyl methanesulfonate (EMS) and subsequent enrichment using growth in D-xylose medium, a strain (TMB3400) was selected that showed over five-fold increase in specific growth rate in D-xylose medium. The study demonstrated the power of random mutagenesis in combination with metabolic engineering to develop superior D-xylose utilizing yeast strains (Wahlbom *et al.,* 2003).

Table 1.2. Strains of *S. cerevisiae* developed by combining metabolic and evolutionary engineering for D-xylose utilization

Strain name	Heterologous pathway expressed	Ethanol yield (g/g D-xylose)	Reference
TMB3001C1	*P. stipitis* XR/XDH	0.24	(Sonderegger and Sauer, 2003)
TMB3400*	*P. stipitis* XR/XDH	0.25	(Wahlbom *et al* 2003)
RWB202-AFX	*Piromyces* XI	0.42	(Kuyper *et al.,* 2004)
RWB218	*Piromyces* XI	0.41	(Kuyper *et al.,* 2005b)
ADAP8	*Orpinomyces* XI	0.48	(Madhavan *et al.,* 2009)
BWY10Xyl*	*C. phytofermentans* XI	0.43	(Brat *et al.,* 2009)
LEK513*	*P. stipitis* XR/XDH	0.31	(Liu and Hu, 2010)
BSPX013	*Piromyces XI*	0.43	(Shen *et al.,* 2012)
H313-A3-AL^{CS}	*Piromyces XI*	0.41	(Zhou *et al.,* 2012)

*industrial strain background

A novel genome shuffling approach has also been introduced to improve the performance of yeast strains for a certain phenotype. In this approach, diploid yeast cells are first mutagenized by EMS to generate more genetic diversity and create useful mutations. The mutant strains are then sporulated and adequately mated with each other to shuffle their genome by virtue of recombination. Selection of the genome-shuffled culture for the desired phenotype resulted in strains with a good combination of alleles and beneficial mutations. Using this method, a strain was developed that showed enhanced ethanol tolerance and produced higher ethanol yield compared to the original strain (Hou, 2009).

More recently, Zhang and Geng, (2012) introduced a modified genome shuffling method to develop a strain of *S. cerevisiae* through recombination of the entire genome of *P. stipitis* with that of *S. cerevisiae* using recursive DNA shuffling. After genomic DNA isolation from *P. stipitis*, and direct transformation into *S. cerevisiae*, a strain that was able to ferment D-xylose to ethanol could be selected.

One of the main disadvantages of evolutionary engineering, which depends on random genetic perturbation and selection for a trait, is that the random genetic alteration might have a negative effect on other important traits. For instance, the first anaerobically growing *S.*

cerevisiae strain in D-xylose medium that was developed by evolutionary engineering for D-xylose fermentation under selective pressure exhibited impaired growth in glucose (Sonderegger and Sauer, 2003). Another limitation attributed to evolutionary engineering is its dependency on selection. Many spontaneous mutations are randomly created and the success of the strategy primarily depends on the effectiveness of the selection method. This means that its application is limited to at most a few phenotypes at a time, such as broad substrate utilization and inhibitor tolerance. Development of efficient selection strategies for phenotypes such as better product formation is a much more challenging task (Nevoigt, 2008).

On the other hand, evolutionary engineering strategies not only help improve strains for certain properties, but also facilitate identification of genetic modifications required for constructing a strain by metabolic engineering. For example, an efficient D-xylose utilizing strain of *S. cerevisiae* was constructed by a combination of metabolic engineering and evolutionary adaptation. The final evolved strain, which showed significantly faster D-xylose assimilation compared to the un-evolved strain was found to have higher expression of XI. This information was then used for further rational engineering of the wild type strain (Zhou *et al.,* 2012). This concept of identifying the genetic basis behind a phenotype of interest and using the information for engineering a superior strain is termed inverse metabolic engineering.

1.5.3. Inverse metabolic engineering

Inverse metabolic engineering has been defined as "the elucidation of a metabolic engineering strategy by first, identifying, constructing, or calculating a desired phenotype; second, determining the genetic or the particular environmental factors conferring that phenotype; and third, endowing that phenotype on another strain or organism by directed genetic or environmental manipulation" (Bailey *et al.,* 2002). In contrast to the rational metabolic engineering approach, inverse metabolic engineering starts with the phenotype but not with the genetic information conferring the phenotype, and therefore is called bottom-up approach.

The major advantage of inverse metabolic engineering is that it does not require prior knowledge about the genetic information underlying the phenotype. As a first step, a strain with the desired phenotype can be identified by screening, or created through evolutionary engineering. The next step is to identify the genetic base establishing the phenotype. As most of the phenotypes of an organism are determined by multiple genetic factors, the identification of the specific genetic change responsible for the phenotype is often challenging. However, the advancement of various molecular biology tools, such as genome sequencing, transcriptomics, proteomics and metabolomics, has facilitated the identification of genetic factors responsible for a certain trait. More recently, the method of pooled-segregant whole-genome sequencing

was shown to be a powerful tool for identification of loci responsible for a quantitative trait (Swinnen *et al.,* 2012a).

Once the genetic basis is identified, the last step is the transfer of the information into the target organism. This is employed by the traditional genetic engineering tool through deletion, over-expression or replacement of the identified gene or regulatory regions. The transfer of the genetic information to a new strain might not fully help achieve the desired phenotype. However, the information obtained can be useful to understand the physiology of the organism. Later on, accumulated information from studies using the inverse metabolic engineering approach will finally lead to a more substantial and effective metabolic engineering strategy. Several successful examples that applied inverse metabolic engineering approaches exist for different applications. Recent examples include the application for increased D-xylose fermentation rate (Zhou *et al.,* 2012), increased ethanol and iso-butanol tolerance (Hong *et al.,* 2010), and improved galactose utilization in *S. cerevisiae* (Hong and Nielsen, 2012).

1.5.4. Global transcription machinery engineering

Global transcription machinery engineering (gTME) is another method of strain improvement employing random genetic alterations. In this approach, key proteins that regulate global transcription are mutated randomly (error prone PCR), which generates new diversity at the transcription level. Clones containing the mutant library are then selected for a certain phenotype. The genetic changes inducing the new phenotype are largely unknown, but, methods like transcriptome analysis can be applied to find causative genes that are up- or down-regulated (Alper *et al.,* 2006; Nevoigt, 2008).

In general the various engineering strategies have their own advantages and disadvantages. Successful application of strain development strategies often relies on the use of a combination of the different engineering strategies. With respect to D-xylose fermentation, all the known metabolic pathways have been expressed in *S. cerevisiae* using rational metabolic engineering. In addition, the defects identified, such as redox imbalance and xylitol accumulation, weak D-xylose uptake rate and slow anaerobic growth could be partially overcome with rational metabolic engineering approaches. Yet, the rate of D-xylose fermentation by even the best recombinant D-xylose utilizing *S. cerevisiae* strain remained very low compared to the glucose fermentation rate. This implies that the current knowledge on the metabolism and regulation of D-xylose assimilation might not be adequate (Cai *et al.,* 2012). However, evolutionary adaptation of such metabolically engineered strains considerably improved the overall ethanol yield and productivity. Since the robust glucose fermentation capability of *S. cerevisiae*, at least of the industrial yeast strains, is the result of long-term human adaptation, the development of strains that utilize D-xylose as well as glucose will not be possible without a better

understanding of the complex physiology, the metabolic pathways and their biochemical regulation. In addition, adaptation of the organism to the conditions in the large-scale industrial environment will play a significant role in developing a robust strain that utilizes D-xylose and glucose with similar rate and efficiency.

1.6. Other advanced biofuel technologies, outlook

1.6.1. Biodiesel

Biodiesel is a *biofuel* that is derived from a mixture of long chain fatty acid alkyl monoesters from vegetable oils and animal fats. Bio-diesel and bioethanol are in fact believed to be the only feasible short-term to medium-term alternative energy sources for the transportation sector. Because of the structural similarity, biodiesel can be used as a replacement of petroleum based diesel. Compared to the fossil diesel, it has numerous advantages. The primary and clear advantage is greenhouse gas savings and its renewability. Other main advantages of biodiesel include; (1) there is no requirement for engine modification since it has similarity to petrodiesel, (2) its high cetane number gives engines a better performance, (3) highly pure bio-diesel eliminates the use of lubricants, (4) combustion of biodiesel produces less soot particles and sulphur (Juan *et al.,* 2011; Su and Wei, 2008).

Biodiesel can be produced by chemical or enzymatic transesterification of triacylglycerols (TAGs) from vegetable oil or animal fat in the presence of alcohol or another acyl acceptor (Figure1.7). Since the use of animal fat is not easily feasible as it forms a solid wax at room temperature, several edible vegetable oils such as palm oil, rapeseed oil, sunflower oil, soybean and corn oil are commonly used. However, as discussed in the section on first generation biofuel production, the use of these biodiesel sources competes with the use for human food consumption. Therefore, non-edible sources, the most promising being jatropha, are getting increasing attention (Juan *et al.,* 2011).

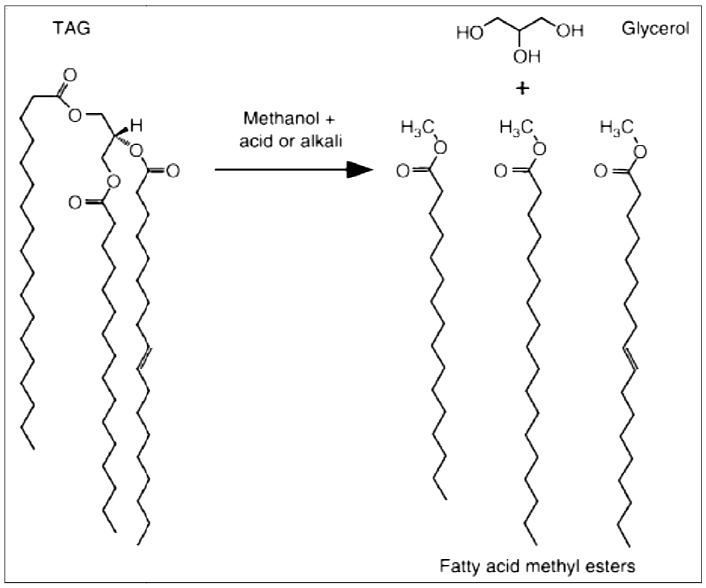


Figure1.7. Esterification of triacylglyceride (TAG) extracted from algae to produce the fatty acid methyl esters (FAMEs) that can be used as biodiesel. In the structure of TAG, glycerol forms an ester bond with three long chain fatty acids. Transesterification of this compound with methanol results in the replacement of the glycerol by methanol, releasing the three FAMEs and glycerol (Scott *et al.,* 2010).

Jatropha is a drought resistant tropical shrub originated from Central America and widely grown in tropical countries like Africa and Asia. The oil content of the seed ranges from 27 to 40% (Achten *et al.,* 2007). A significant advantage of using this plant as an oil source for biodiesel is its ability to grow on marginal lands without competing with food crops. In addition, the seeds mature in only 3 to 4 months after flowering and once it grows to an adult plant, it persists in producing seeds for about 50 years (Juan *et al.,* 2011).

Transesterification of jatropha or other vegetable oils is mainly performed though chemical or enzymatic catalysis. The common commercial process employs alkaline catalysis to convert vegetable oil or animal fat to fatty acid methyl esters (FAME) in the presence of methanol. Though chemical catalysis achieves higher conversion of the TAGs into the corresponding esters, this method is energy intensive, requires wastewater treatment, and it is difficult to remove the glycerol by-product and the catalyst from the product. The use of lipase enzymes is more attractive since they are more efficient, require less energy and produce less waste. On the other hand, enzyme conversion is not economically viable due to the high cost of the enzymes and their short operational activity because of their inactivation by the alcohol present in the reaction and the glycerol produced (Akoh *et al.,* 2007). Various studies have been conducted to address these problems, especially the inhibition of the enzyme by alcohol. One approach is to add methanol stepwise rather than adding a high molar concentration at once. This approach significantly stimulated the transesterification and gave a higher product yield (Shimada *et al.,* 2002). Although addition of solvent to reduce enzyme inactivation is also

effective, the method is less applicable since it generates toxic waste products. In addition, the purification step necessary to remove the solvent from the product adds additional cost to the overall production process. In another approach, the use of alkyl acetate instead of methanol as an acyl acceptor was shown to significantly reduce the inactivation of enzymes. In a study of transesterification of jatropha oil in the presence of ethyl acetate instead of methanol, the enzyme activity could be well maintained even over 12 cycles of enzyme reuse (Modi *et al.,* 2007).

Overall, jatropha oil is believed to be one of the future renewable sources for advanced biodiesel production. Chemical conversion of the oil into fatty acid alkyl esters is effective but the high cost of the downstream process and its environmental effects limits the use of this method at commercial scale. Transesterification using lipases is a good alternative but further work is required to produce the enzymes in a cost effective way and to optimize the conditions in order to obtain a commercially viable process.

1.6.2. *Biofuels* from algae

Algae are a group of aquatic eukaryotes ranging from unicellular (microalgae) to multicellular (macroalgae) organisms. Microalgae have great potential to be the future feedstock for biodiesel production. This is because of their fast growth rate (some species have a doubling time of only a few h), their ability to accumulate high concentrations of TAGs (that can be converted to biodiesel), and their ability to grow in a wide variety of water sources including sea water and brackish water, and in areas that are unsuitable for agriculture (Chisti, 2008; Scott *et al.,* 2010).

Efficient production of biodiesel from microalgae depends on the species of algae used and the growth conditions. In general, algae produce different amounts of carbohydrates, lipids and proteins depending on the species and growth conditions. The lipid content produced by specific species also depends on the specific substrate and environmental conditions, and can vary between 20 to 80% per dry weight. High concentrations of lipids in microalgae can be induced by environmental stress (Menetrez, 2012).

Growth of microalgae can be conducted either in open ponds or in photobioreactors. In an open field growth system, algae are usually grown in natural waters like lakes and ponds or artificial ponds or containers, which is usually shallow. Although open ponds are easy to operate and are cheap, they have several drawbacks including difficulty of proper control of the growth conditions, and high risk of contamination from bacteria and wild algae. The use of closed plastic or glass tubular photobioreactors allows proper management, in order to obtain pure high-oil microalgae. Nutrient water and the algae culture itself can be circulated in the

bioreactor to keep the microalgae from settling and to expose them to natural sun light, which is the source of energy for their lipid production (González-Fernández *et al.,* 2011; Lehr and Posten, 2009).

Compared to agricultural oil seeds, biodiesel from algae is believed to be able to completely replace petrodiesel without significantly affecting the supply of food and other agricultural products (Chisti, 2008). However, one of the most significant challenges to obtain economically feasible biodiesel production from algae is the cost of downstream processing, starting from biomass harvesting and extraction of the oil. When the algae biomass grows enough, harvesting can be accomplished though various ways including filtration, centrifugation and flocculation. Once the biomass is separated, the extraction of the oil from the biomass is commonly employed by mechanical pressing and chemical solvents. All these processes are too expensive to be economically viable. In addition to optimizing the growth and extraction system for the algae, maximizing the TAG production will have significant importance in reducing the overall cost of downstream processing (Scott *et al.,* 2010).

1.6.3. Biomass-to-liquids (BtL) diesel (Fischer-Tropsch diesel)

BtL diesel is one of the methods for production of clean biofuels from biomass. Its production involves initial conversion of biomass into synthetic gas (syngas) such as hydrogen (H_2), carbon monoxide (CO), carbon dioxide (CO_2), and methane (CH_4). After a cleaning process, the synthetic gas, specifically CO and H_2, are catalytically converted in a Fischer-Tropsch process (hydrogenation of CO using metallic catalysts to produce long chain hydrocarbons) into liquid hydrocarbons such as biodiesel, biokerosine and jet fuel. Among the biggest challenges in biomass-to-liquid fuel conversion is the cost of generation of syngas and its purification. These processes alone can account up to 75% of the capital cost, indicating that improvement of these steps is critical for commercial scale application of the process (van Steen and Claeys, 2008).

To sum up, the environmental merit associated with *biofuel*s is very large on condition that they are produced efficiently. Biofuels have many other advantages, including energy security, foreign currency savings because of reduced imports, as well as overall contribution to economic development. For these reasons, the biofuel economic sector is believed to grow significantly over the coming years. Though it is difficult to forecast which biofuel will dominate in the future, bioethanol and bio-diesel will probably become the dominant short-term to medium-term alternatives to fossil fuels in the transportation sector.

Chapter 2

Development of an efficient D-xylose fermenting industrial *S. cerevisiae* strain for advanced bioethanol production

Adapted from:
Demeke MM, Dietz H, Li Y, Foulquié-Moreno M, Mutturi S, Deprez S, Den Abt T, Bonini B , Liden G, Dumortier F, Verplaetse A, Boles E and Thevelein JM. **"Development of a D-xylose fermenting and inhibitor tolerant industrial yeast strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering**". *Biotechnology for Biofuels, 2013, **6**:89*

Author's contribution
The author participated in the design of the project, the experimental work, data analysis and manuscript preparation.

2.1 Introduction

The engineering of novel metabolic capacities into robust microorganisms may be easier than the alternative strategy, i.e. the engineering of very high ethanol tolerance and prominent general robustness into strains, which have the desired metabolic capacity. Impressive progress has been made in engineering pentose fermentation capacity into the yeast *S. cerevisiae* (Hahn-Hägerdal *et al.,* 2007; Nevoigt, 2008). For that purpose, two heterologous pathways for D-xylose utilization have been employed. First, the genes encoding D-xylose reductase (XR) and xylitol dehydrogenase (XDH) from *Pichia stipitis* have been expressed in *S. cerevisiae*. This resulted in D-xylose fermentation, but also in significant production of xylitol under anaerobic conditions, which is due to NADH/NADPH cofactor imbalance of XR and XDH (Hahn-Hägerdal *et al.,* 2007b). The performance of these strains has been improved considerably by addressing the cofactor imbalance and by over-expression of endogenous xylulokinase (XK) and enzymes of the non-oxidative part of the pentose phosphate pathway (Bengtsson *et al.,* 2009; Olofsson *et al.,* 2011; Runquist *et al.,* 2010).

The second pathway allows direct isomerization of D-xylose to xylulose through heterologous expression of xylose isomerase (XI). After the first successful attempt to express the thermophilic bacterium *Thermus thermophilus* XI into *S. cerevisiae* (Walfridsson *et al.,* 1996)**,** recombinant strains expressing the fungal *Piromyces sp.* Strain E2 xylose isomerase have been developed, which displayed higher enzymatic activity (Kuyper *et al.,* 2005a, 2003b). By the use of an isomerization instead of a reduction/oxidation conversion of D-xylose to xylulose, the potential problem of co-factor imbalance is avoided. However, the rate of D-xylose utilization in XI expressing strains was initially found to be inferior to that in strains harboring the XR/XDH pathway (Bettiga *et al.,* 2008). This was mostly attributed to the low activity of the XI enzyme in *S. cerevisiae* and its inhibition by xylitol, generated from reduction of D-xylose by the endogenous enzyme aldose reductase encoded by *GRE3*. The level of xylitol produced is much lower, however, than in the strains expressing the XR/XDH pathway. Knocking out *GRE3* in an XI expressing strain improved both the rate of D-xylose consumption and ethanol production (Traff *et al.,* 2001). However, the aldose reductase, encoded by *GRE3,* plays a role in stress protection and its deletion is therefore not desirable in industrial yeast strains (Aguilera and Prieto, 2001). To overcome these problems, Brat *et al.* (2009) constructed the first recombinant *S. cerevisiae* strain with high activity of prokaryotic XI, using codon optimized *xylA* gene from *Clostridium phytofermentans.* This enzyme was much less inhibited by xylitol compared to the enzyme from *Piromyces*. Nevertheless, the rate of D-xylose consumption and ethanol production by this recombinant strain was still low.

Different metabolic and evolutionary engineering strategies have been used successfully to improve D-xylose utilization in a yeast strain expressing *Piromyces* xylose isomerase.

Overexpression of genes encoding xylulokinase and enzymes of the non-oxidative part of the pentose phosphate pathway, combined with deletion of *GRE3* to reduce xylitol formation, considerably improved the D-xylose consumption rate (Kuyper *et al.,* 2005a). Combining metabolic engineering with evolutionary adaptation resulted in strains with strong pentose fermentation capacity and partial cofermentation of glucose, D-xylose and even L-arabinose (Kuyper *et al.,* 2005b; Wisselink *et al.,* 2009). These results suggested that the xylose isomerase pathway might be the pathway of choice for constructing superior industrial yeast strains with optimal fermentation performance in lignocellulose hydrolysates (Van Maris *et al.,* 2007). However, all these evolutionary engineered strains were still made in a haploid laboratory yeast strain background, displaying in general suboptimal fermentation performance and poor robustness and stress tolerance, which make them unsuitable for use in industrial fermentations. Since previous work showed that XI expressing strains displayed higher yield of ethanol per consumed D-xylose compared to strains harboring the XR/XDH pathway (Bettiga *et al.,* 2008), and since they profit from direct isomerization of D-xylose to xylulose without cofactor requirement, the XI pathway seemed to be most promising to engineer into a robust industrial yeast strain.

In this work, we have selected Ethanol Red as industrial host strain to engineer high-capacity D-xylose fermentation, because Ethanol Red is one of the most widely used yeast strains for first-generation bioethanol production (Fermentis, a division of S. I. Lesaffre, Lille, http://www.fermentis.com/). The strain has excellent glucose fermentation capacity, high robustness and stress tolerance, and also displays excellent performance in fed-batch production on molasses, is tolerant to dehydration and retains high vitality during storage and transport. Using this strain, we aim to develop an industrial *S. cerevisiae* strain that converts D-xylose to ethanol with a yield close to the theoretical maximum yield and with a very high specific rate of fermentation. For that purpose, a D-xylose- and L-arabinose-utilization gene cassette has been stably integrated into both alleles of the *PYK2* locus in the diploid industrial bioethanol production strain Ethanol Red (by Prof. Eckhard Boles' group, Goethe University of Frankfurt, Germany). The gene cassette contains genes coding for xylose isomerase (XI) from *Clostridium phytofermentans,* endogenous xylulokinase (XK) and all the enzymes of the non-oxidative part of the pentose phosphate pathway; transketolase (*TKL1*), transaldolase (*TAL1*), ribulose-5-phosphate 3-epimerase (*RPE1*) and ribose-5-phosphate ketol-isomerase (*RKI1*). The arabinose metabolism genes include codon-optimized genes for arabinose transporter (AraT) from *Pichia stipitis*, optimized arabinose isomerase (AraA) from *Bacillus licheniformis*, optimized ribulokinase (AraB) and ribulose-5-P 4-epimerase (AraD) from *E. coli,* as well as the optimized endogenous Transketolase 2 (*TKL2*) and Transaldolase 2 (*NQM1*). The resulting recombinant strain was called HDY.GUF5. Despite the presence of both gene cassettes in the genome, the strain could hardly ferment any D-xylose or L-arabinose to ethanol. In this study, we used Ethyl

Methanesulfonate (EMS) mutagenesis, genome shuffling and selection in lignocellulose hydrolysate enriched with D-xylose, and subsequent evolutionary adaptation in YP + D-xylose, to greatly enhance D-xylose utilization efficiency and maintain the inhibitor tolerance and general robustness of the background strain.

2.2 Evolutionary engineering of the recombinant industrial yeast strain HDY.GUF5 for D-xylose fermentation

2.2.1 Overview of the evolutionary engineering strategy

The evolutionary engineering strategy used in this study is shown in figure 2.1. To render high D-xylose fermentation capacity in the recombinant industrial strain HDY.GUF5, we focused first on D-xylose fermentation performance, since D-xylose is the dominant pentose sugar in lignocellulose hydrolysates, and since it was unclear whether L-arabinose utilization might compromise D-xylose utilization in this strain background. The HDY.GUF5 strain was subjected to several consecutive processes for strain improvement: mutagenesis, genome shuffling and selection, followed by multiple rounds of evolutionary adaptation (Figure 2.1).

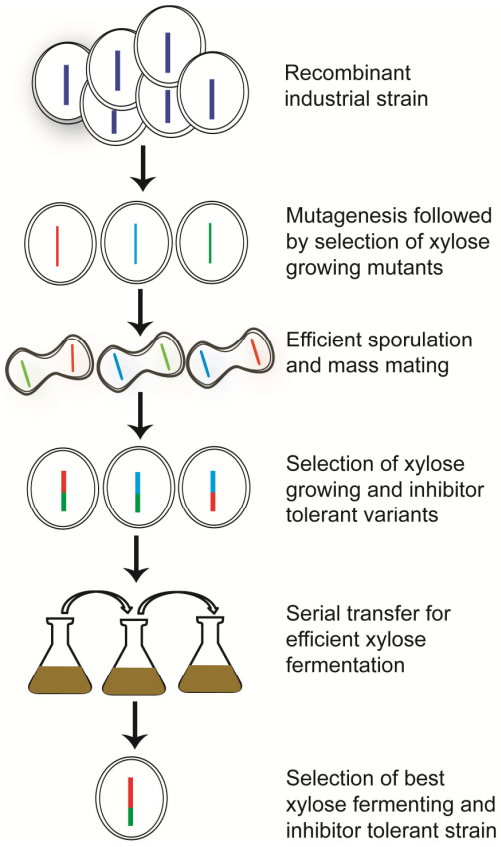


Figure 2.1. General strategy used for development of a strain with high D-xylose fermentation capacity and high inhibitor tolerance. The diploid recombinant industrial strain carrying both the D-xylose and L-arabinose cassette was mutagenized and mutant strains able to grow on D-xylose were selected on YPX. The genomes of the selected mutants and the parent strain HDY.GUF5 were shuffled by sporulation and mass mating, and D-xylose utilizing cells were selected in D-xylose-enriched pre-treated spruce hydrolysate. The culture was then submitted to

evolutionary adaptation in YP + 40 g/L D-xylose, single clones were evaluated in different stages and the strain with the best performance was selected from the last stage.

2.2.2 EMS mutagenesis of recombinant strain HDY.GUF5

The diploid HDY-GUF5 strain was treated with 3% EMS for 0.5, 1, 2, 3, and 4 h in order to create diverse populations of mutant strains with possibly beneficial mutations for D-xylose utilization. The mutagenized cells were plated out onto YPX plates, and also on YPD plates to estimate the cells' survival rate. The rate of survival after EMS mutagenesis decreased significantly with increased incubation time (Figure 2.2). After 4 h of EMS treatment, only 3.3% of cells were able to survive. In medium with D-xylose as a sole carbon source, 4 colonies from the 4 h treated and 3 colonies from the 3 h treated cells were obtained after incubation for 72 h. No colonies were recovered in YPX plates from the EMS untreated cells and those treated for 0.5 to 2 h.

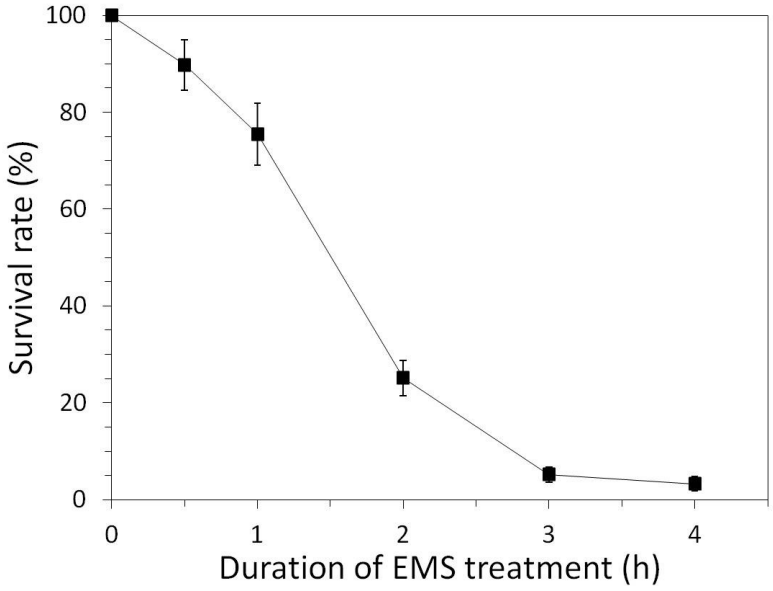


Figure 2.2. Survival percentage of yeast cells treated for different durations with 3 % EMS. The relative number of EMS treated cells growing in YPD plate relative to the EMS untreated cells was expressed as survival rate. Error bars represent SD from duplicate experiments.

The seven D-xylose-growing strains were plated for single colonies and further tested in semi-anaerobic batch fermentations with D-xylose. The rate of D-xylose fermentation by all mutants was extremely low (data not shown), except for mutant M315, which showed slightly better fermentation compared to the parent strain HDY-GUF5 (Figure 2.3 a). It also displayed a two times higher growth rate (0.037 h^{-1} ± 0.004) compared to the parent strain (0.019 h^{-1} ± 0.0004) in synthetic medium containing D-xylose as sole carbon source.

2.2.3 Genome shuffling

The sporulation efficiency of the seven mutants was evaluated before the application of genome shuffling. Only one of the mutants, M492, was still able to sporulate. The 3 to 4 h EMS treatment possibly caused mutations abolishing sporulation capacity in the other strains. The best D-xylose utilizing mutant strain M315 was found to have *MATα* mating type and was able to mate with *MAT**a*** cells. Therefore, the two mutant strains, M315 and M492, together with the parent strain HDY.GUF5 were used for the genome shuffling step. The parent strain was included in order to facilitate loss of deleterious mutations.

The M492 mutant strain and the parent strain HDY.GUF5 were sporulated to more than 75% efficiency and the spores liberated by zymolyase treatment. The mass of isolated spores from the two strains was allowed to germinate in YPD and then mass-mated with exponentially growing cells of the mutant M315. The zygotes were allowed to proliferate at 35°C in synthetic medium containing D-xylose as sole carbon source. The OD_{600} increased from 2.5 to 12 in 24h. The whole cell population was then transferred into undetoxified spruce hydrolysate, supplemented with YP and 4% D-xylose. The concentration of spruce hydrolysate used prevented growth of the parent strain HDY.GUF5, but the shuffled culture was able to grow in 48 h from an initial OD_{600} of 2 to an OD_{600} of 26. Undetoxified acid-pretreated spruce hydrolysate supplemented with 40 g/L D-xylose was chosen for selection, because it contains a high amount of inhibitors and only a limited amount of D-glucose (13 g/L). When the D-glucose was used up, the strains continued to grow on the supplemented D-xylose, allowing us to select inhibitor tolerant mutants without losing the capacity to grow on D-xylose.

In order to test the spore viability as a validation of the genome shuffling step, samples obtained after spore purification from both HDY-GUF5 and M492 were spread for single colonies on YPD plates. Only a few colonies (less than 0.1%) from M492 were able to germinate, while the expected number of colonies from the wild type strain HDY.GUF5 was able to grow. We then dissected 9 tetrads from each strain. While none of the spores from M492 tested were viable, the expected number of cells germinated from the parent HDY.GUF5, indicating that mainly the HDY.GUF5 and the mutant M315 were involved in the genome shuffling step, while the M492 strain likely had a much lower or no contribution at all.

2.2.4 Directed evolution

In order to enrich for fast D-xylose utilizing clones and subsequently improve the D-xylose utilization rate, the entire population of cells obtained after genome shuffling and subsequent selection in spruce hydrolysate with D-xylose, was used for the evolutionary engineering process. The cells were first grown aerobically in shake flasks containing 40 ml YPX medium for 48 h and then used for inoculation at an initial OD_{600} of 2.75 into cylindrical 150 ml

fermentation tubes containing 100 ml YP medium with 40 g/L D-xylose. The fermentations were performed at 35°C under semi-anaerobic conditions, which were attained by slow stirring of the culture at 120 rpm to insure mixing of the cells without significant aeration. This method gradually created semi-anaerobic conditions (oxygen level of less than 1 ppm) within 1 h of incubation. The fermentation rate of the first culture, called GS1.1 was slow but already much better compared to the original strains used in the genome shuffling (Figure 2.3a). We then performed ten additional serial transfers using the same medium, each time with an initial cell density with OD_{600} of 5 (equivalent to 1.3 g DW/L). A relatively high inoculation density was used to insure that new variants of the cell population that were generated during the evolutionary process were effectively transferred to the next batch. In addition, complex medium, rather than defined mineral medium, was chosen for cultivation to avoid selective pressure due to nutrient limitation. As a result, the D-xylose utilization rate was the main selective criterion.

In the second culture, GS1.2, the lag phase was much longer than in the first culture, GS1.1, probably due to loss of viability during the prolonged incubation of the GS1.1 culture. However, a sharp rise in D-xylose consumption rate, as indicated by the CO_2 production rate, was observed after 112 h (Figure 2.3a). To avoid possible loss of viability because of substrate depletion, subsequent serial transfers were performed each time before complete D-xylose depletion. Considerable improvement in the rate of D-xylose consumption was observed with each round of evolutionary adaptation (Figure 2.3a). The most dramatic change happened in the 3rd culture, GS1.3, in which almost no lag phase was observed, as opposed to the 112 h lag phase in the previous GS1.2 culture. In the 8th, 9th and 10th culture, the concentration of D-xylose was increased to 50 g/L, 60 g/L and 100 g/L, respectively, to further adapt the yeast to higher D-xylose concentrations with the assumption that higher concentrations of D-xylose might increase the rate of fermentation due to higher flux through the pathway. For the last culture, GS1.11, 40 g/L D-xylose was used again to make sure that the strain could also utilize lower D-xylose concentrations at a similar rate.

The progress of the evolutionary engineering process was continuously monitored by evaluation of single cell clones. A total of 9, 15, 20 and 27 single cell isolates from the 2nd, 4th, 7th and 11th culture, respectively, that were able to grow well on YPX plates, were evaluated for fermentation performance in YP + 40 g/L D-xylose (Figure 2.3b). A previously constructed industrial D-xylose utilizing strain, TMB3400, expressing xylose reductase and xylitol dehydrogenase (Wahlbom *et al.,* 2003), and the parent strain HDY.GUF5, were included for comparison (Figure 2.3b). Some isolates from the 2nd and 4th culture were already better than TMB3400 in terms of D-xylose fermentation rate. However, all isolates from the 7th and 11th culture showed a much faster rate and much higher extent of D-xylose utilization than

TMB3400 (Figure 2.3b). The individual clones isolated from the 7th and 11th culture showed a similar rate and extent of fermentation and therefore the evolutionary adaptation process was terminated after the 11th culture.

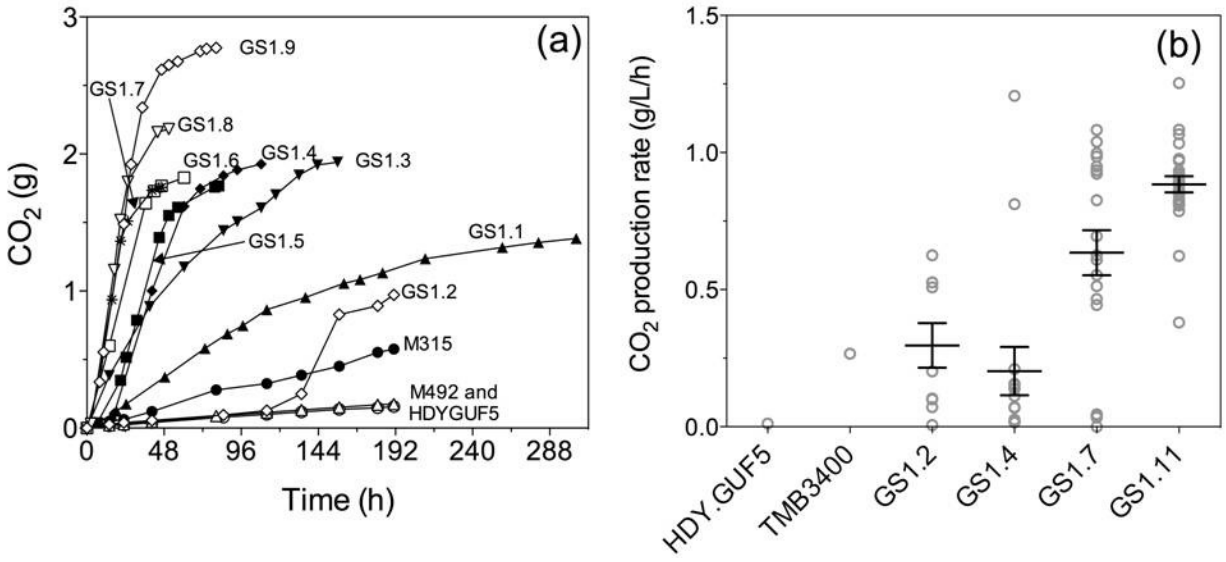


Figure 2.3. Gradual establishment of efficient D-xylose utilization during the mutagenesis and multiple evolutionary engineering steps. (a) CO_2 production as measured by weight loss in sequential semi-anaerobic batch fermentations in YP + 4% D-xylose at 35°C. After EMS mutagenesis and one step of genome shuffling, the culture was submitted to 11 serial transfers, in which each time part of the culture after the batch fermentation was used to start a new fermentation. The CO_2 production profile of the first 9 serial batch fermentations is shown. The concentration of D-xylose was increased to 5% and 6% in the 8th and 9th batch fermentation, respectively. (GS1 stands for the one step of genome shuffling, and the next number indicates the step in the serial transfer). (b) Volumetric CO_2 production in semi-anaerobic batch fermentation in YP + 4% D-xylose at 35°C by single cell isolates obtained from the 2nd (GS1.2), 4th (GS1.4), 7th (GS1.7) and 11th (GS1.11) serial batch fermentation during the evolutionary adaptation process. The horizontal bar represents the mean with the standard error of the mean. The parent strain, HDY.GUF5, and a previously constructed industrial D-xylose utilizing stain, TMB3400, [28] are shown for comparison.

2.3 Selection of the best D-xylose fermenting strain

Seven of the best individual clones from the 4th, 7th and 11th culture were tested in more detail for fermentation performance and ethanol yield in YP medium with 40 g/L D-xylose. The course of fermentation with the best strains from the cultures GS1.4, GS1.7 and GS1.11, as well as the control strain HDY.GUF5, is shown in Figure 2.4a. The final ethanol level for these fermentations is shown in Figure 2.4b. The seven isolates showed a similar performance with a slight difference in the rate of CO_2 production and final ethanol yield. Ethanol yield of up to 0.48 g/g was obtained for most of the strains, with 40 g/L D-xylose as the main carbon source and an initial cell density of 1.3 g DW/L. This corresponds to 94% of the theoretical maximum ethanol

production. There was almost no xylitol and little glycerol produced. Although the final ethanol level reached by these strains was very similar, the isolate GS1.11-26 reproducibly showed the highest rate of fermentation (Figure 2.4a) and was selected for further characterization.

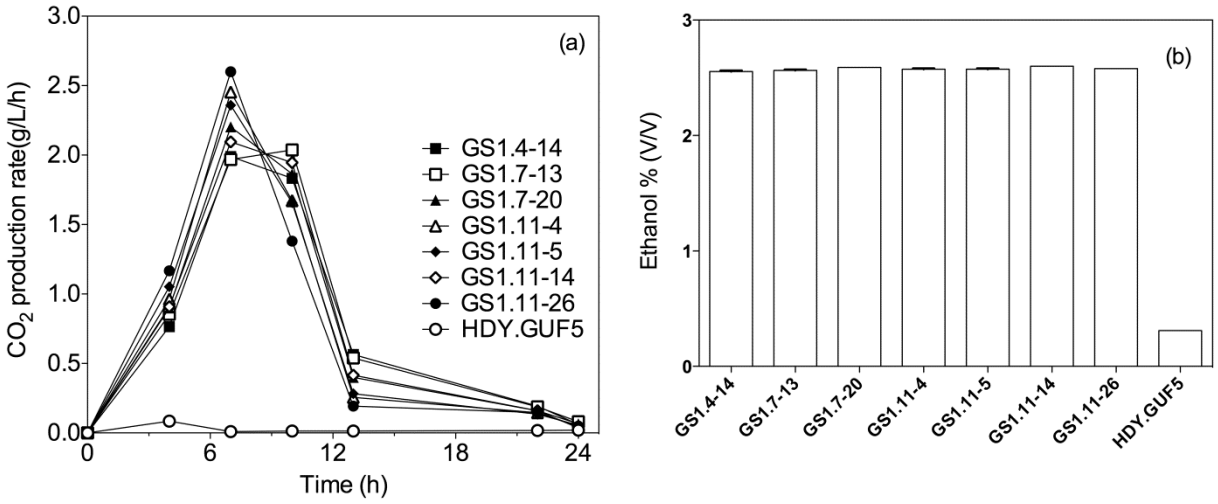


Figure 2.4. D-xylose fermentation by superior single-cell isolates from different serial batch fermentations in the evolutionary engineering procedure. (a) CO_2 production as measured by weight loss in semi-anaerobic batch fermentations in YP + 4% D-xylose at 35°C. Selected single cell isolates from each of the 4th (GS1.4), 7th (GS1.7) and 11th (GS1.11) culture were tested. The parent strain, HDY.GUF5 was used for comparison. The average of a duplicate experiment is shown. (b) Final ethanol titer reached in the fermentations of (a). Error bars represent the standard error from duplicate experiments.

We also tested mating type and ploidy of the best performing single cell clones from the 11th culture, GS1.11. The GS1.11-26 strain and all other strains tested, as well as the mutant M315, were found to be *MATα* and to have a diploid DNA content (Figure 2.5 and results not shown). We also performed a pheromone assay for mating type to rule out the possibility that the mating type PCR failed due to SNPs in the HMR locus, which might cause failure to detect *MAT**α***. However, the pheromone assay confirmed that the strains were *MATα* (results not shown). Hence, we can conclude that the best D-xylose utilizing clones were *MATα/α* diploids.

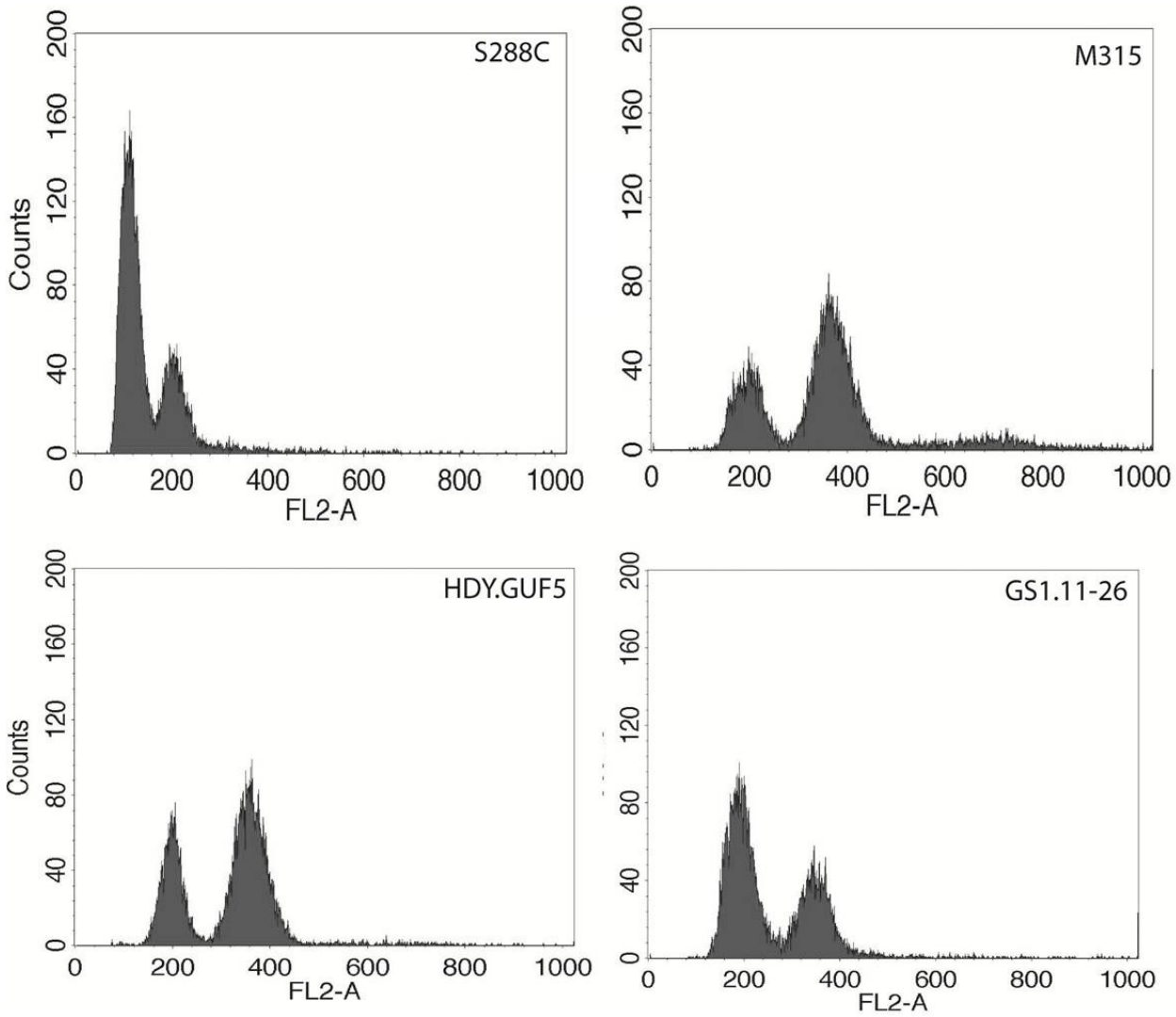


Figure 2.5. Comparison of DNA content among parent and mutant strains, as determined by flow cytometry. DNA content is shown for a haploid control strain S288c, mutant M315, diploid parent strain HDY.GUF5 and strain GS1.11-26, the best performing single cell isolate obtained from the 11th culture. Mutant M315 and GS1.11-26 are shown to be diploid.

2.4 Fermentation performance of strain GS1.11-26 with D-xylose and a glucose/D-xylose mixture

The fermentation performance of the strain GS1.11-26 was evaluated in semi-anaerobic batch fermentation at 35°C with an initial cell density of 1.3 g DW/L. Synthetic medium or YP medium with D-xylose or a mixture of D-xylose and glucose was used for the evaluation (Figure 2.6). An air-tight fermentation lock containing glycerol was used to avoid entrance of air. Samples were taken every few h with needles.

In synthetic medium with 35 g/L D-xylose as sole carbon source, the evolved strain consumed all the available D-xylose in about 17 h, (Figure 2.6a) with maximum D-xylose consumption rate

of 1.10 g/g DW/h and maximum ethanol production rate of 0.48 g/g DW/h. The final yield of ethanol was 0.46 g/g D-xylose and the xylitol yield was less than 0.01 g/g D-xylose. Though yeast strains expressing XI produce less xylitol compared to those of XR/XDH expressing strains (Kuyper *et al.,* 2003b), the presence of other enzymes such as the non-specific aldose reductase encoded by *GRE3* results in conversion of D-xylose to xylitol (Kuhn *et al.,* 1995). Since the xylitol yield in the strain GS1.11-26 was very low, we sequenced the *GRE3* gene in both the parent and the final evolved strain to test for possible mutations abolishing its function. However, the sequences were found to be identical in both strains.

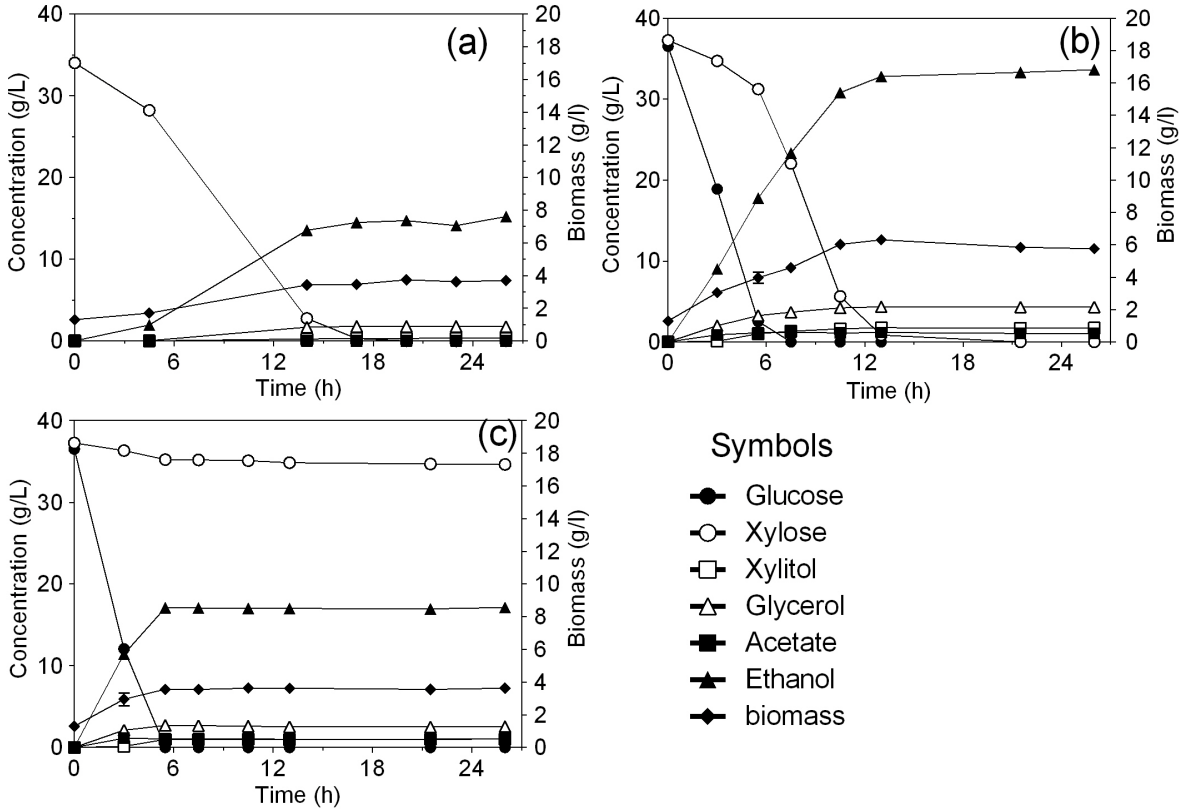


Figure 2.6. Performance of strain GS1.11-26 in semi-anaerobic batch fermentations with D-xylose and a glucose/D-xylose mixture. (a) Semi-anaerobic batch fermentation in synthetic medium with D-xylose with strain GS1.11-26. (b) Semi-anaerobic batch fermentation in rich YP medium containing 36 g/L glucose and 37 g/L D-xylose with strain GS1.11-26. (c) Semi-anaerobic batch fermentation in rich YP medium containing 36 g/L glucose and 37 g/L D-xylose with parent strain HDY.GUF5.

Co-fermentation in rich YP medium containing 36 g/L glucose and 37 g/L D-xylose was used to compare the performance of the evolved strain, GS1.11-26, with that of the parent strain HDY.GUF5 (Figure 2.6 b and c). In this condition, both glucose and D-xylose were almost completely consumed in about 13 h by the evolved strain, with maximum ethanol productivity of 1.4 g/g DW/h. Compared to the parental strain, the evolved strain showed an 8.5-fold faster rate of D-xylose consumption: 1.10 versus 0.13 g/g DW/h (Table 2.1). However, the maximum

specific glucose consumption rate was slightly higher in the parent strain (1.4-fold) (Table 2.1). Although the overall ethanol yield per consumed sugar was the same in both strains, the ethanol yield per initial sugar present in the medium was about 2-fold higher for the evolved strain compared to the parent strain. This is due to the fact that after 32 h only 5% of the D-xylose was consumed by the parent strain. In the evolved strain, D-xylose consumption started from the beginning of the fermentation (Figure 2.6b) but remained slow as long as there was glucose present (the first 5 h). The co-consumption of D-xylose during glucose fermentation occurred at a rate of 0.4 g/g DW/h. Afterwards, D-xylose consumption strongly accelerated and attained a volumetric rate only slightly lower than that of glucose consumption. However, the biomass also increased significantly during the glucose fermentation period, so that more biomass was present during the period of fastest D-xylose consumption compared to the period of fastest glucose consumption. As a result, the maximum specific glucose consumption rate (2.71 ± 0.04 g/g DW/h) was about 2.5 times faster than the maximum specific D-xylose consumption rate (1.10 ± 0.00 g/g DW/h), which was reached just after glucose was completely exhausted (Table 2.1).

Table 2.1 Comparison of fermentation performance between the parent strain, HDY.GUF5, and the evolved strain, GS1.11-26, in YP medium containing 36 g/L glucose and 37 g/L D-xylose.

Strain	Maximum specific sugar consumption rate (g/g DW/h)		Yield (g/g sugar) [a]			Specific ethanol productivity (g/g DW/h)
	Glucose	D-Xylose	Ethanol	Xylitol	Glycerol	
HDY.GUF5	3.83±0.08	0.13±0.01	0.23±0.00	0.03±0.00	0.04±0.00	1.79±0.08
GS1.11-26	2.71±0.04	1.10±0.00	0.46±0.00	0.04±0.00	0.06±0.00	1.38±0.01

Note: [a]: Yield of ethanol and glycerol were calculated per g total sugar while yield of xylitol was calculated per g D-xylose.

The co-utilization of glucose and D-xylose was also evaluated at a higher glucose concentration to understand the effect of higher glucose concentration on D-xylose utilization and to see if D-xylose can be co-utilized with glucose (Figure 2.7). In YP medium containing a mixture of 147 g/L glucose and 43 g/L D-xylose, the D-xylose utilization rate was much slower in comparison to the rate obtained during fermentation where the glucose concentration was lower (Figure 2.7 a). However, slight D-xylose consumption was observed even before the glucose was completely exhausted. Nevertheless, about 90% of the D-xylose was consumed in 36 h. After glucose was completely exhausted, the D-xylose utilization rate did not return to a faster state like the rate observed during the fermentation in the presence of lower glucose concentration or in the absence of glucose. This might be due to the inhibitory effect of the high concentration of ethanol already produced from the glucose fermentation phase.

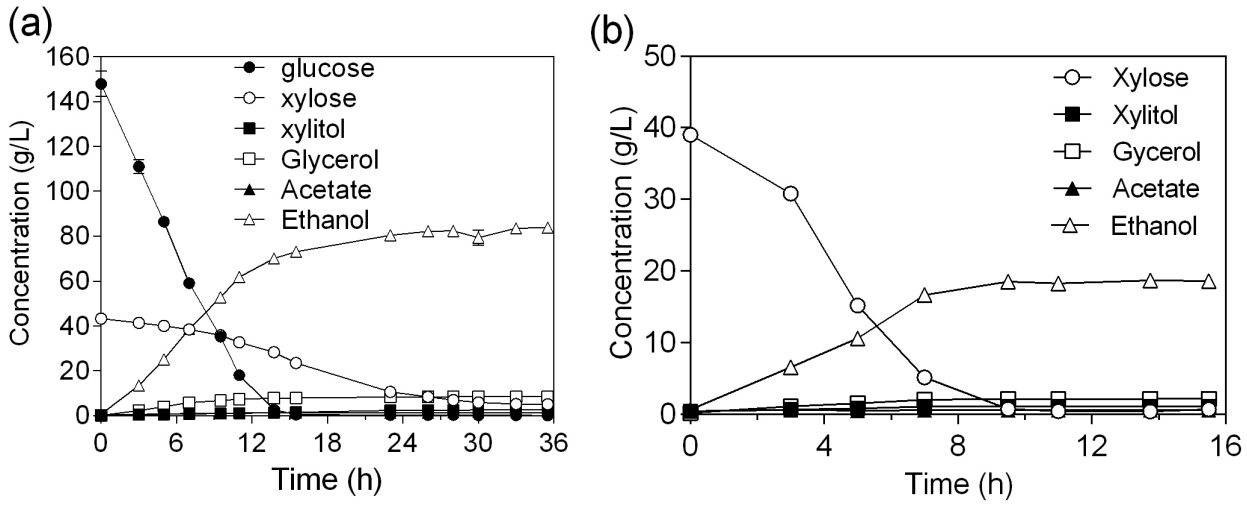


Figure 2.7 Semi-anaerobic batch fermentation by GS1.11-26 in YP medium containing a mixture of 150 g/L glucose and 40 g/L D-xylose (a), and 40 g/L D-xylose alone (b). Data shown is an average from duplicate experiments and error bars represent standard errors of the mean from the two duplicates.

Since the D-xylose utilization rate was lower in the beginning of the fermentation even in the absence of glucose, we performed an experiment in which we started the fermentation first with D-xylose alone, and subsequently injected glucose once rapid D-xylose utilization had been established. In this condition, the rate of D-xylose consumption was severely impaired until the injected glucose was consumed (data not shown). This indicates an inhibitory effect of glucose on D-xylose utilization during D-xylose/glucose co-fermentation. As *S. cerevisiae* does not encode xylose specific transporters, it has to use hexose transporters that have much higher affinity for glucose than for D-xylose (Hamacher *et al.,* 2002). The manifestation of a decreased D-xylose utilization rate immediately after addition of glucose might therefore be due to competitive inhibition of D-xylose transport by glucose.

2.5 Stability of the D-xylose fermentation phenotype in strain GS1.11-26

We have assessed the stability of the D-xylose fermentation phenotype through many generations of growth in rich glucose medium in the absence of any D-xylose. For that purpose, three independent colonies of strain GS1.11-26 were inoculated into 5 ml YPD and serially transferred, at 100 times dilution each, for about 50 generations. Subsequently, replicate samples were spread onto YPD and YPX plates and the ratio of the number of colonies growing on YPX relative to that on YPD was calculated.

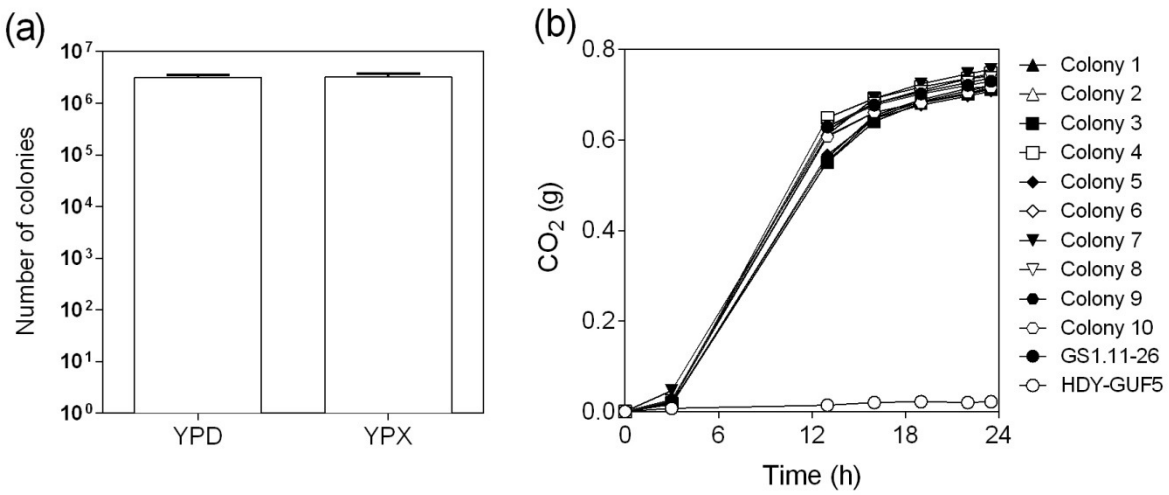


Figure 2.8. Evaluation of the stability of D-xylose fermentation capacity in strain GS1.11-26. (a) Comparison of the number of colonies growing in YPD and YPX after growth for about 50 generations in non-selective conditions with only glucose as carbon source. Error bars represent standard error of the mean from three independent experiments. (b) Fermentation performance in YP + 4% D-xylose as estimated from weight loss due to CO_2 release. Colonies 1 to 10 represent randomly selected single-cell clones isolated after growth of GS1.11-26 for 50 generations in YPD medium. The parent strain HDY.GUF5 is shown for comparison.

As shown in figure 2.8, the number of colonies on the YPD and YPX plates was very similar. Moreover, there was no apparent difference in the size of the colonies growing on the YPX plates. When ten randomly selected single-cell clones, obtained from the last YPD culture, were tested for fermentation performance in YPX, all ten showed a fermentation performance very similar to that of the original GS1.11-26 strain (Figure 2.8 b). These results indicate that the D-xylose fermentation capacity of the GS1.11-26 strain was completely stable in the absence of any selection pressure.

2.6 Discussion

The XI from *C. phytofermentans* was the first prokaryotic XI that showed high activity upon expression in *S. cerevisiae,* both in laboratory and industrial strains (Brat *et al.,* 2009). However, the industrial strain expressing the codon-optimized version of the gene could only ferment D-xylose to ethanol after further evolutionary adaptation in D-xylose medium. Though the rate of D-xylose utilization by the evolved strain was much too low to allow industrially viable ethanol production from lignocellulosic feedstocks, the work provided a starting point for the development of strains using a bacterial XI that was less inhibited by xylitol than the fungal *Piromyces* XI. In the present work we used systematic evolutionary engineering, and developed a robust industrial *S. cerevisiae* strain that efficiently converts D-xylose to ethanol with high yield and productivity.

Rational metabolic engineering alone was not able to establish efficient D-xylose (or L-arabinose) utilization capacity in the Ethanol Red strain. The recombinant strain HDY.GUF5 failed to show D-xylose (or L-arabinose) fermentation. Efficient rational engineering strategies rely on the complete understanding of the metabolic network, as well as its regulation in response to the dynamic environmental conditions to which the engineered strain is exposed (Nevoigt, 2008). Because of the complexity and still limited understanding of the biological and regulatory network of D-xylose metabolism in recombinant *S. cerevisiae* strains, rational approaches have faced huge challenges to eliminate the factors that limit efficient D-xylose fermentation (Chu and Lee, 2007). Several of these factors have been identified (Chu and Lee, 2007; van Maris *et al.,* 2007). Most of these requirements have been addressed in the strain HDY.GUF5, which include overexpression of the PPP genes, *XKS1* and the hexose/pentose transporter encoding gene *HXT7*, as well as codon optimization of some of the genes based on the highly efficient glycolytic codon usage of yeast (Wiedemann and Boles, 2008). Although expression of the same codon-optimized XI in a laboratory strain established moderate D-xylose fermentation, the industrial strain used in this study as well as previously (Brat *et al.,* 2009), was not able to metabolize D-xylose. This is likely due to the difference in the genetic background of the strains, although the precise mechanism remains unclear (Hector *et al.,* 2010).

Combining metabolic engineering with evolutionary adaptation alone or together with random mutagenesis has been proven successful for developing strains with improved D-xylose fermentation efficiency (Kuyper *et al.,* 2005a; Liu and Hu, 2010; Sonderegger and Sauer, 2003; Wisselink *et al.,* 2009; Zhou *et al.,* 2012). In addition, genome shuffling has also been used in combination with metabolic engineering and evolutionary adaptation, for improving D-xylose utilization capacity in different *S. cerevisiae* strains (Jingping *et al.,* 2012; Zhang and Geng, 2012). In the present paper, we successfully exploited a combinatorial approach using all three

random strain improvement strategies described above, in order to improve D-xylose fermentation efficiency of the recombinant industrial strain HDY.GUF5.

We first started with random mutagenesis of the recombinant strain to generate very diverse genetic variation that might establish initial D-xylose fermentation capacity. Selection of mutants with a significant anaerobic D-xylose fermentation rate is a challenging task, because likely multiple mutations are required (Nevoigt, 2008). In addition, a previous study reported that direct selection of a mutant *S. cerevisiae* population capable of anaerobic D-xylose utilization was unsuccessful (Sonderegger and Sauer, 2003). Therefore, we first selected clones from a heavily-mutagenized population that were able to grow at least to some extent on D-xylose medium as a sole carbon source. Since strong random mutagenesis likely results in both beneficial and deleterious mutations, we presumed that genetic recombination of the mutants obtained, with the original industrial strain by genome shuffling, and selection for D-xylose utilization capacity, would result in enrichment of beneficial and loss of unfavourable mutations. After only one step of genome shuffling, the whole shuffled culture already demonstrated a significantly improved rate of D-xylose fermentation. However, attempts to isolate single cell clones from this shuffled culture with better D-xylose utilization rate, compared to that of the best mutant strain M315, failed. Thus, we decided to enrich the clones with most rapid D-xylose utilization and at the same time further improve their rate of D-xylose utilization, through adaptive evolution in D-xylose medium. The selection of clones with a shorter lag phase on D-xylose and a higher D-xylose utilization rate is most obvious. However, clones that utilize a larger part of the D-xylose will be able to undergo more proliferation cycles and therefore will tend to be present in higher amounts and thus also preferentially transferred to the next culture. Since the cultures were semi-anaerobic, the D-xylose is largely converted to ethanol and therefore these clones will likely also have a higher ethanol yield.

After only two transfers in D-xylose medium under semi-anaerobic conditions, the D-xylose fermentation rate already increased dramatically. Subsequent serial transfers resulted in further gradual improvement of D-xylose utilization. In previous studies, evolutionary adaptation under aerobic condition followed by gradual transition to anaerobic condition, was necessary to obtain strains with anaerobic D-xylose utilization capacity. In addition, several generations were required to obtain strains with efficient D-xylose fermentation capacity (Kuyper *et al.,* 2005b, 2004; Sonderegger and Sauer, 2003). In our study, even though the best isolate GS1.11-26 was isolated after 11 serial transfers, clones with high D-xylose utilization capacity could already be isolated after only 2 transfers. The rapid improvement in the rate of D-xylose fermentation might be explained by the presence of a suitable combination of important genetic changes introduced by the mutagenesis and genome shuffling, and sustaining rapid improvement of D-xylose utilization by a repetitive subsequent genetic

modification, such as amplification of the *xylA* gene or another crucial genetic element, and/or rapid enrichment of clones with a superior combination of mutant alleles. An important genetic change might have been generated also in the second culture of the evolutionary adaptation, which was characterized by a sharp rise in CO_2 evolution at the end, and a dramatic increase in the rate of fermentation when this culture was transferred to the next batch. During evolutionary engineering, expansions and contractions of different subpopulations can occur (Almario *et al.,* 2013a; Reyes *et al.,* 2012) and an individual cell with a beneficial mutation, providing a relative fitness advantage, can develop into a dominant subpopulation after several generations in serial batch transfer experiments (Elena and Lenski, 2003). In our work, high variability in the rate of D-xylose fermentation was observed among individual clones isolated from intermediate cultures in the evolutionary adaptation process. However, isolates from the last culture showed a very similar fermentation performance, although not precisely the same, suggesting that the fitter clones finally conquered and dominated the culture.

The best strain, GS1.11-26, showed a reproducible and stable D-xylose fermentation rate and was further characterized in synthetic or complex medium containing only D-xylose or mixture of glucose and D-xylose. In synthetic medium with D-xylose as a sole carbon source, the GS1.11-26 strain showed a maximum specific D-xylose consumption rate at least 15 times higher than the previous industrial strain BWY10Xyl expressing the same codon-optimized *C. phytofermentans* XI (Brat *et al.,* 2009). The GS1.11-26 strain also accomplished complete attenuation of D-xylose with an ethanol yield of 0.46 g/g D-xylose, whereas the previous strain BWY10Xyl left a substantial amount of D-xylose unfermented. Moreover, the yield of ethanol obtained with GS1.11-26 was higher than the yield obtained with the best strain reported recently (Zhou *et al.,* 2012). As a consequence, GS1.11-26 exhibited the highest D-xylose to ethanol conversion yield than any other recombinant strain of *S. cerevisiae* reported so far. The high ethanol yield can also be explained by the very low xylitol yield, which is remarkable since the *GRE3* gene had not been deleted nor was it inactivated in the strain development programme. The low xylitol yield, in the absence of *GRE3* inactivation, might be due to the inherently higher metabolic flux in the industrial bioethanol production strain Ethanol Red, compared to the previously used strain backgrounds.

In conclusion, we demonstrated that combining rational metabolic engineering with a systematic approach of random strain modification can be used as an efficient means for development of strains with efficient D-xylose fermentation. The best strain obtained in this work, GS1.11-26, has highly promising potential for use and further development of an all-round robust yeast strain for efficient fermentation of various lignocellulose hydrolysates. Moreover, it already contains the genes for additional utilization of L-arabinose and should be easily evolved also for efficient fermentation of this pentose sugar.

Chapter 3

Evaluation of the efficient D-xylose fermenting strain GS1.11-26 for industrially relevant traits

Adapted from:

Demeke MM, Dietz H, Li Y, Foulquié-Moreno M, Mutturi S, Deprez S, Den Abt T, Bonini B, Liden G, Dumortier F, Verplaetse A, Boles E and Thevelein JM. **"Development of a D-xylose fermenting and inhibitor tolerant industrial yeast strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering**". *Biotechnology for Biofuels, 2013, **6**:89*

Author's contribution

The author participated in the design of the project, the experimental work, data analysis and manuscript preparation.

3.1 Introduction

Economically viable lignocellulose based ethanol production requires organisms that are able to ferment all sugars, including pentose sugars, present in the lignocellulose hydrolysates. In addition, the ideal organism should also exhibit significant inhibitor tolerance in order to be able to ferment the sugars in non-detoxified lignocellulose hydrolysates. Considerable progress has been made in the past few years to develop yeast strains that are able to ferment D-xylose (Hahn-Hägerdal *et al.,* 2007b; Nevoigt, 2008), and strains with improved inhibitor tolerance (Gorsich *et al.,* 2006; Larsson *et al.,* 2001b; Larsson *et al.,* 2001a). However, combining both inhibitor tolerance and broad substrate utilization remains a challenge. Few D-xylose fermenting recombinant strains of *S. cerevisiae* and natural D-xylose utilizing yeast strains with improved inhibitor tolerance have been reported (Hou and Yao, 2012; Koppram *et al.,* 2012). However, both the D-xylose fermentation capacity and inhibitor tolerance have to be improved further, in order to use them for fermentation of lignocellulosic biomass at a higher solid loading. High solid loading is one of the feasible methods for industrial ethanol production with second generation biomass (Olofsson *et al.,* 2008a).

In this respect, the development of the D-xylose fermenting strain GS1.11-26 from a prime bioethanol production yeast Ethanol Red has promising potential for use of the strain as an all-round robust yeast strain for efficient fermentation of various types of lignocellulosic biomass. Ethanol Red has a proven record of excellent fermentation capacity, high robustness and stress tolerance, and also displays excellent performance in fed-batch production on molasses, is tolerant to dehydration and retains high vitality during storage and transport. However, the random mutagenesis and evolutionary adaptation steps used to develop the strain GS1.11-26 might have resulted in accumulation of several mutations that might affect other industrially important traits. Hence, we evaluated the performance of this strain for several industrially relevant traits, such as fermentation capacity in real industrial substrates, ethanol accumulation capacity and rate of fermentation under very high gravity fermentation conditions, as well as aerobic growth rate in glucose medium.

3.2 Evaluation of GS1.11-26 for industrially relevant traits

3.2.1 Fermentation performance of strain GS1.11-26 in lignocellulose hydrolysates

3.2.1.1 Separate hydrolysis and fermentation (SHF)

SHF was performed to evaluate the fermentation performance of the strain GS1.11-26 in hydrolysates of three industrially relevant feedstocks: *Arundo donax* (giant reed), *Picea abies* (Norway spruce), and a 50/50 mixture of wheat straw and hay (Figure 3.1). Pretreated materials of giant reed and spruce were hydrolyzed using the enzyme cocktail ACCELLERASE® 1500 at 53 °C for 48 h. Pretreated material of the wheat straw/hay mixture was hydrolyzed with Novozymes cellulase complex and β-glucosidase at 50 °C for 24 h (KaHo Sint-Lieven, Ghent,

Belgium). The whole slurry obtained after hydrolysis was used to start the fermentation. Yeast extract (10 g/L) and peptone (20 g/L) were added as a source of nitrogen, vitamins, amino acids and other nutrients.

In all the three hydrolysates tested, the strain GS1.11-26 produced a high yield of ethanol, close to the theoretical maximum (Table 3.1). The highest yield of ethanol was obtained with the wheat straw/hay hydrolysate, reaching 0.48 g/g of initial glucose and D-xylose, which is equivalent to 94% of the maximum theoretical yield. In this hydrolysate, no xylitol and only small amounts of glycerol and acetate were detectable at the end of the fermentation, which might have contributed to the high yield.

Table 3.1. Composition of the three lignocellulose hydrolysates and fermentation yield of the evolved strain, GS1.11-26.

Medium	Initial sugar (g/L)			Initial inhibitor concentration (g/L)			Yield (g/g sugars)		
	Glucose	D-xylose	Mannose	Acetate	HMF	Furfural	Ethanol	Xylitol	Glycerol
Arundo donax	59.34	22.18	6.19	6.80	0.32	0.14	0.47±0.01	0.07±0.00	0.04±0.00
Spruce	48.91	10.09	14.97	4.79	1.09	1.57	0.43±0.00	0.003±0.000	0.04±0.00
Wheat straw/hay	68.22	28.22	ND	0.70	ND	ND	0.48±0.02	ND	0.04±0.00

ND: Not detectable

Arundo donax hydrolysate contained the highest acetate concentration from the three hydrolysates tested (table 3.1). An acetate concentration of about 4 g/L is known to be inhibitory to growth and fermentation of most yeast strains at low pH (Zhang *et al.,* 2011; Bellido *et al.,* 2011). Moreover, D-xylose fermentation is more sensitive to acetate inhibition (Bellissimi *et al.,* 2009; Wright *et al.,* 2011). Despite the presence of an initial acetate concentration of 6.8 g/L in the *Arundo donax* hydrolysate used, the strain consumed all the glucose and more than 90% of the D-xylose in about 96 h with an ethanol yield of 0.47 g/g total sugar, equivalent to 92% of the maximum theoretical ethanol yield (Figure 3.1a and Table 3.1). A final ethanol titer of 5.1 % (v/v) was reached in 96 h.

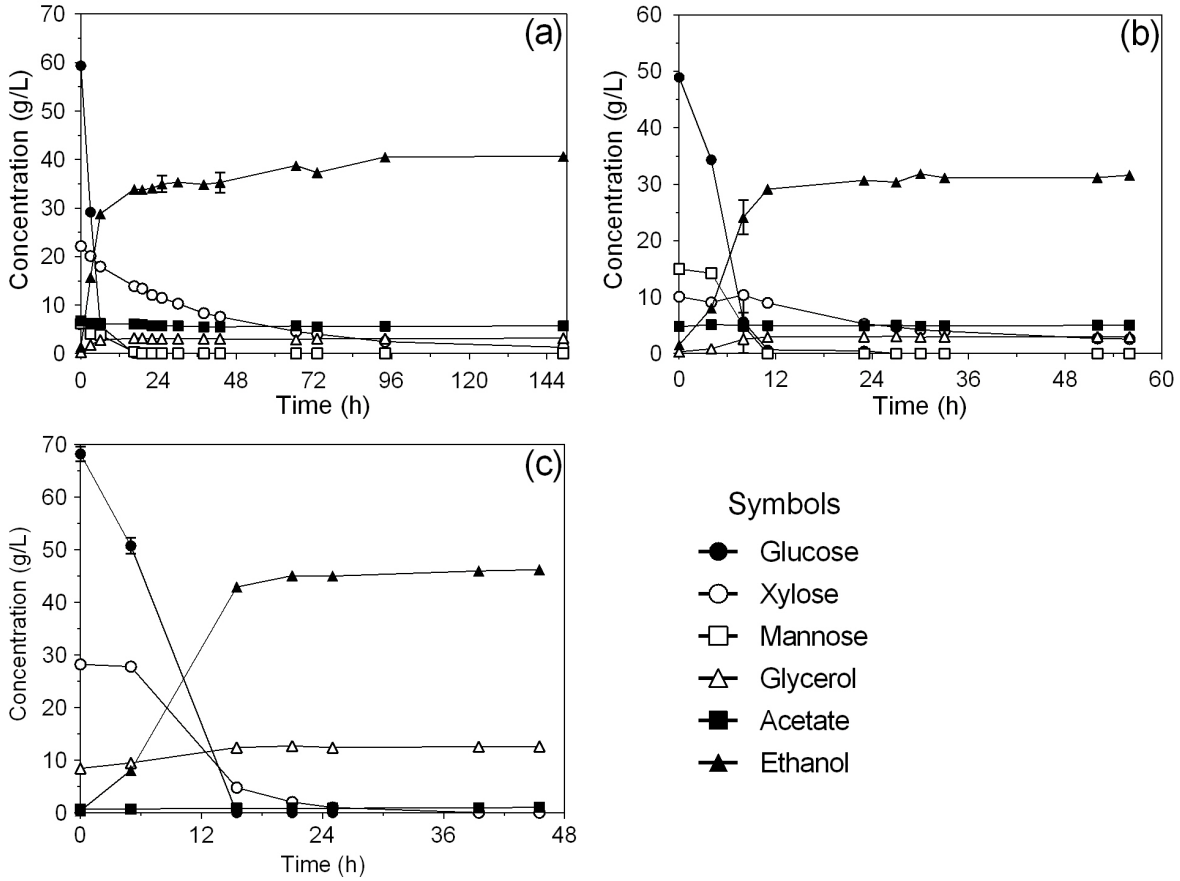


Figure 3.1. Performance of strain GS1.11-26 in semi-anaerobic batch fermentations with three different lignocellulose hydrolysates. (a) *Arundo donax* (giant reed), (b) spruce and (c) mixture of wheat straw/hay.

Acid pretreated spruce has been reported to be among the most inhibitory hydrolysates, since it contains high concentrations of phenolic compounds, weak acids and furan derivatives, that synergistically inhibit yeast growth and fermentation (Shuai *et al.,* 2010; Jönsson *et al.,* 2013). Although we measured only the furans and acetic acid, the concentration of these compounds, especially HMF and furfural, was indeed much higher in spruce than in the other two hydrolysates (table 3.1). Since the amount of D-xylose in spruce was very low (about 10 g/L), this highly inhibitory spruce hydrolysate was used mostly to evaluate the inhibitor tolerance of the strain GS1.11-26. Accordingly, furfural and HMF were completely consumed in 4 h, while acetate remained in the medium. In spite of the elevated inhibitor concentrations, a high ethanol yield of 0.43g/g initial sugars was produced. This is remarkable in view of the high inhibitor concentrations in spruce hydrolysate. A final ethanol titer of 4% (v/v) was reached already in 30 h.

The wheat straw/hay hydrolysate contained the highest initial sugar concentration and the lowest level of inhibitors (Table 3.1). Both glucose and D-xylose were completely consumed in about 24 h producing a final ethanol concentration of 45.07± 0.92 g/L (Figure 3.1 c). This was

the highest ethanol concentration of all three hydrolysates. The rapid fermentation and complete sugar attenuation in wheat straw/hay hydrolysate is likely due to the low level of inhibitors (Table 3.1).

The yield of ethanol per g consumed glucose and D-xylose was higher in all the lignocellulose hydrolysates compared to that in synthetic and YP medium (chapter 2). This is probably due to the lower amount of xylitol and glycerol formed, and is consistent with previous results (Olofsson *et al.,* 2008b)

Next, we evaluated the tolerance of the GS1.11-26 strain, in comparison with its parent strain HDY.GUF5, to individual inhibitors: HMF, furfural and acetic acid. The strains were inoculated in synthetic medium with glucose and pH 4.5, in the presence of a range of concentrations of the inhibitors. In this condition, the evolved strain GS1.11-26 showed similar tolerance to HMF and furfural, showing only partial inhibition up to 10 g/L HMF and 5 g/L furfural. However, the tolerance of the evolved strain to acetic acid was reduced (normal growth up to 5 g/L acetate as measured after 72 h) compared to the parent strain (normal growth up to 7 g/L acetate as measured after 72 h) (Figure 3.2).

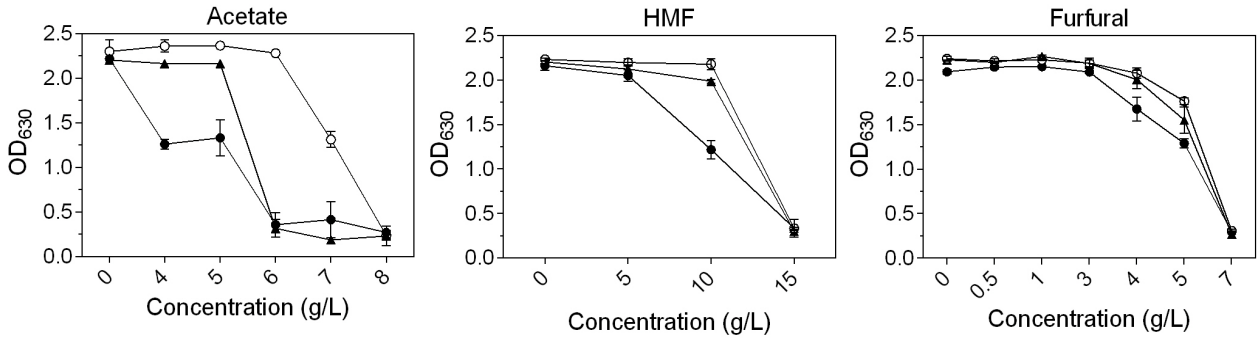


Figure 3.2: Evaluation of strain GS1.11-26 for tolerance to HMF, furfural and acetic acid. Growth assay was performed in 24-well plates containing 1 ml of synthetic complete medium with 20 g/L glucose and various concentrations of each inhibitor. Strains were inoculated at an initial OD_{630} of 0.2. Error bars represent standard error of the mean from duplicate experiments. Strains: (●) GS1.11-26, (○) parent HDY.GUF5, (▲) M315.

3.2.1.2 Simultaneous saccharification and fermentation (SSF).

The performance of the evolved strain was also tested in an SSF using both pretreated Arundo and spruce material (performed by Prof. Gunnar Liden's group, Lund University, Sweden). In a previous study, the yields obtained for SHF and SSF with pretreated Arundo were compared using the XR/XDH strain VTT C-10880 (Ask *et al.,* 2012). In that study, the overall yield in SHF was higher than that obtained in SSF. The main reason was most likely that SSF was performed at a temperature of 32 ^{o}C, which resulted in a low degree of enzymatic hydrolysis. This was confirmed in a study in which SSF was run at both 32 and 39 ^{o}C with the strain Ethanol Red (Mutturi and Lidén, 2013). For this reason, it was decided to assess the strain GS1.11-26 in SSF at 39^{o}C with Arundo hydrolysate and to compare with the results

for Ethanol Red reported in the previous study (Mutturi and Lidén, 2013). Due to the favorable D-xylose to glucose ratio (Olofsson *et al.,* 2008b), D-xylose was consumed already from the beginning of the SSF, and remained below 1 g/L from 24 h to 96 h, while the xylitol formation was negligible (Figure 3.3).

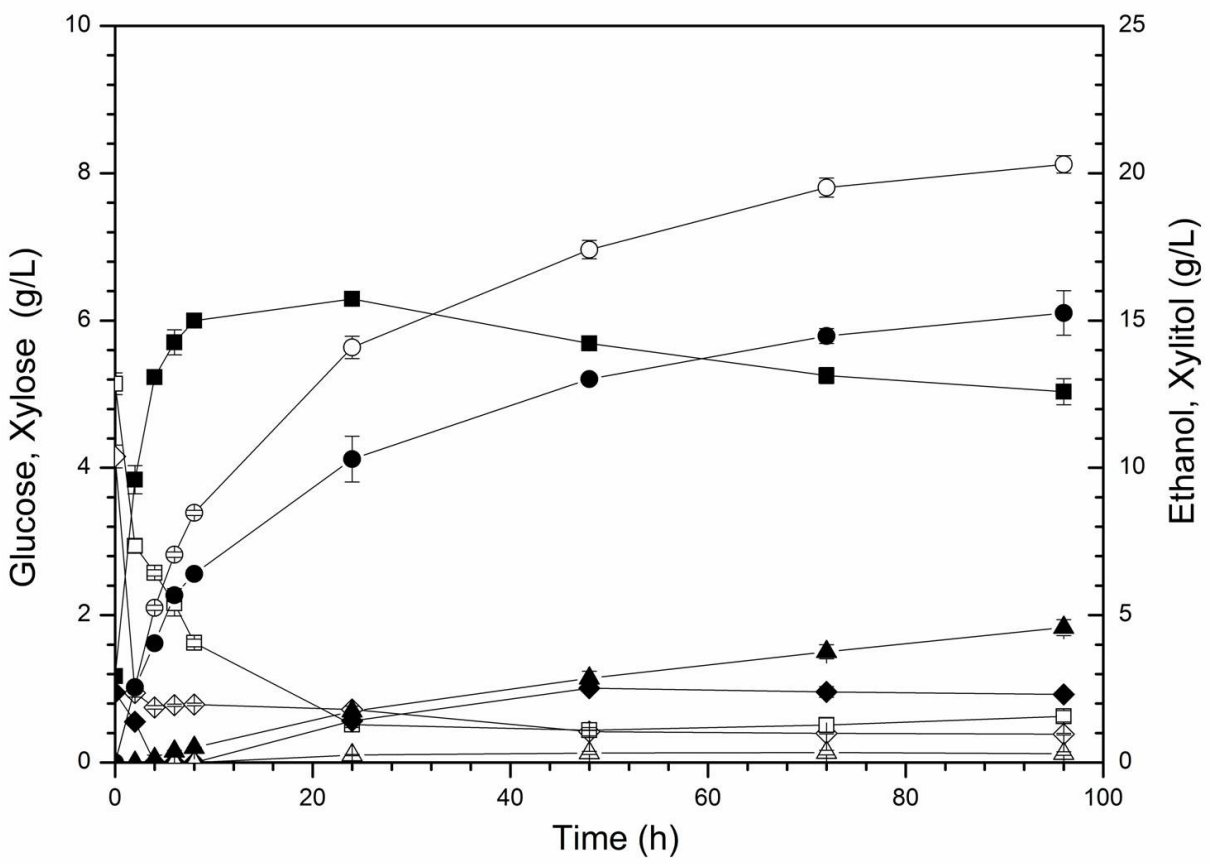


Figure 3.3. Performance of strain GS1.11-26 in SSF with pretreated Arundo. Medium composition during SSF of pretreated Arundo (10% WIS) using GS1.11-26 (open markers). As a comparison, previously reported SSF values (Mutturi and Lidén, 2013) with the strain Ethanol Red (filled markers) are shown. Symbols used: glucose (diamond), ethanol (circle), xylose (square) and xylitol (triangle). A yeast concentration of 4 g DW/L was used and a temperature of 39°C. The enzymes used were Celluclast and Novozyme 188.

The strain GS1.11-26 was also tested in SSF using pretreated spruce. Although the xylan content of spruce is much lower than that of Arundo, the spruce material provided a suitable test of inhibitor tolerance, and SSF experiments were made at both 32 and 39 ^{o}C. D-xylose consumption was more efficient at 32 ^{o}C with less than 1 g/L of D-xylose and xylitol at the end of 96 h in the case of spruce (cf. Table 3.1). The fermentation rate was reduced after 48 h at 39^{o}C (data not shown), and higher amounts of residual glucose and D-xylose were found after 96 h (Table 4). The optimum temperature for SSF was thus lower for spruce than for Arundo.

Table 3.2: Final sugar and metabolite concentrations measured after 96 h SSF with the strain GS1.11-26 using steam pretreated spruce. A WIS content of 10% was used in the experiments.

	SSF Temperature	
	32 °C	39 °C
Residual glucose (g/L)	0.79	3.69
Residual xylose (g/L)	0.40	2.26
Xylitol (g/L)	0.70	0.35
Final ethanol concentration	35.3	30.6
Yield (g ethanol/g total sugar)*	0.32	0.28
% of maximum yield	63.9	55.7

*Includes only fermentable sugars from fiber and liquid fractions

3.2.2 Performance of strain GS1.11-26 in very high-gravity fermentations

The strain GS1.11-26 was also tested for tolerance to high osmolarity and high ethanol in very high gravity (VHG) fermentations. This is a challenging quality test for industrial yeast, because it requires a combination of very high osmotolerance and very high ethanol tolerance (Liu *et al.,* 2011; Pais *et al.,* 2013). The original, untransformed parent strain Ethanol Red had previously shown very good performance in unstirred, semi-anaerobic VHG fermentations, accumulating in YP medium about 17-18% (v/v) ethanol from 330 g/L glucose (Pais *et al.,* 2013). The same performance was observed in the genetically modified parent strain, HDY.GUF5 (Figure 3.4). On the other hand, the evolved strain, GS1.11-26, showed a slower fermentation rate and a 1-2% (v/v) lower final ethanol titer under the same VHG conditions (Figure 3.4). This indicates that adverse background mutations have been introduced in the evolved strain during the EMS treatment or that they were spontaneously generated during the evolutionary engineering. An alternative explanation is that the beneficial mutations that improve D-xylose fermentation have adverse side-effects on other properties, like osmotolerance and/or high ethanol tolerance.

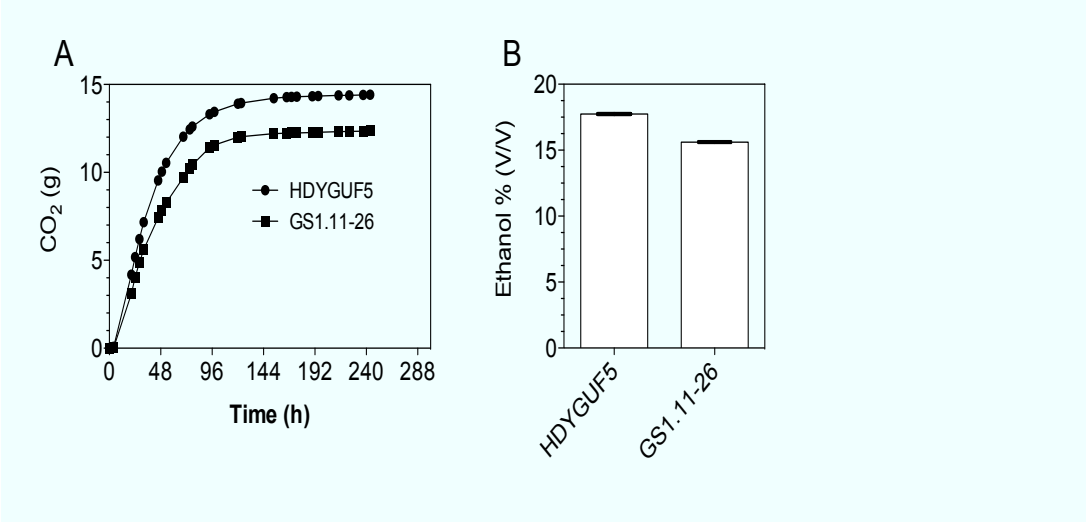


Figure 3.4. Performance of strain GS1.11-26 as compared to HDY.GUF5 in very-high gravity fermentation. (a) CO_2 production as measured by weight loss in semi-anaerobic batch fermentations in YP + 33% glucose at 30 °C. (b) Final ethanol titer reached in the fermentations of (a).

To evaluate whether the slow fermentation performance in VHG fermentation is due to impaired tolerance to osmotic or ethanol stress, we examined the tolerance of the evolved strain to both stresses in comparison to the parent strain. Figure 3.5 shows the results of a growth test by dilution spot assays on solid YPD medium containing a high concentration of sorbitol (for osmotic stress) or a high concentration of ethanol. Though the tolerance to osmotic stress by the evolved strain GS1.11-26 was similar to that of the parent strain HDY.GUF5, the ethanol tolerance was severely reduced in the evolved strain, which was manifested by moderate growth at an ethanol concentration of only 14% (v/v). The parent strain was able to grow at an ethanol concentration of up to 17% (v/v).

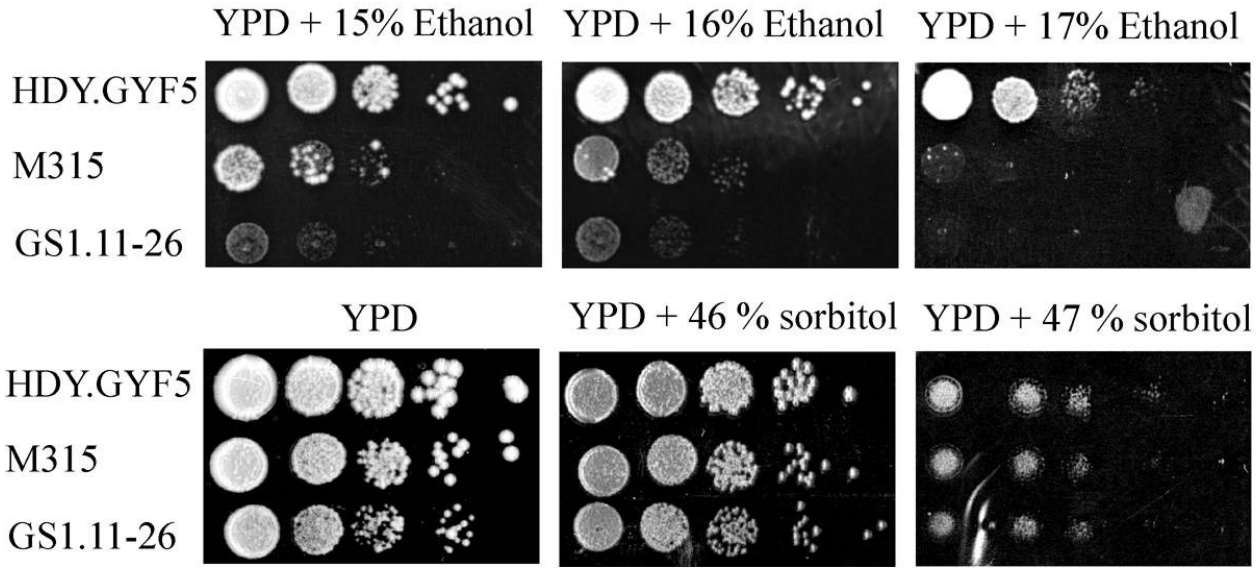


Figure 3.5. Effect of ethanol and osmotic stress on growth of GS1.11-26. Growth on YPD plates containing different concentrations of ethanol or sorbitol was performed with a spot assay in 10-fold serial dilutions from an initial OD_{600} value of 0.5. The assay was performed two times with similar results using independent cultures.

3.2.3 Growth rate of strain GS1.11-26 under aerobic conditions

The production of a maximal amount of yeast cell biomass under aerobic conditions is one of the requirements for industrial yeast propagation. The latter is performed in highly-controlled aerobic fed-batch fermentation in which the sugar level is maintained at a very low level to avoid any production of ethanol. Hence, the yeast should be able to combine a high growth rate with a purely respiratory metabolism to maximize the production of biomass (Bellissimi, 2009). With this notion, the strain GS1.11-26 was evaluated for growth rate under aerobic conditions in different media and volumes. In a bioscreen assay with synthetic medium containing 20 g/L glucose, the maximum respiro-fermentative growth rate of the evolved strain, GS1.11-26, was only about 75% (0.342 ± 0.005 h^{-1}) of that of the parent strain, HDY.GUF5, (0.459 ± 0.021).

Moreover, the two mutants M315 and M492, derived from HDY.GUF5, grew faster than GS1.11-26 and slower than HDY.GUF5 (Figure 3.6a). This might indicate that the genetic changes causing the slower aerobic growth rate in GS1.11-26 occurred both during mutagenesis and the subsequent genome shuffling and/or evolutionary adaptation process.

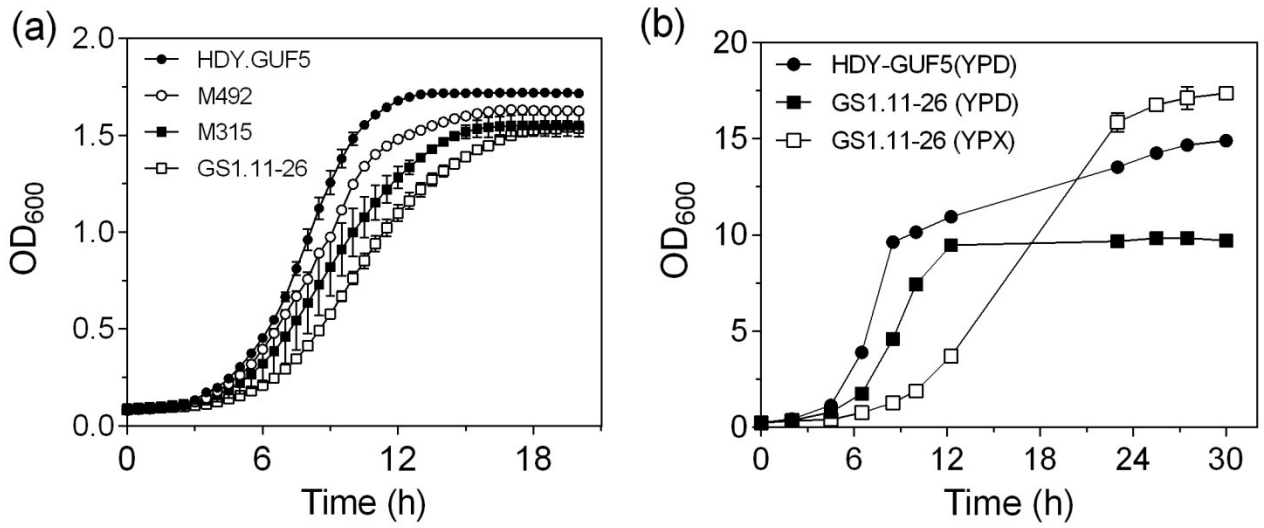


Figure 3.6. Growth rate of GS1.11-26 (a) growth rate in synthetic medium containing 2% D-xylose as a sole carbon source in a bioscreen assay in 200 µl volume at 35 °C. (b) growth rate in 40 ml YP medium containing either D-xylose and glucose. Error bars represent standard error of the mean of triplicate (bioscreen assay) or duplicate (growth assay in 40 ml culture) experiments.

In shake flask cultures with complex medium (40ml YPD), a similar difference between the evolved and parent strain was observed. The aerobic respiro-fermentative growth rate of the evolved strain, GS1.11-26, was only 80% (0.485± 0.032 h^{-1}) of that of the parent strain, HDY.GUF5 (0.614 ± 0.096 h^{-1}). Moreover, there was no second growth phase after the diauxic shift in the evolved strain (Figure 3.6 b). Hence, the evolved strain was apparently unable to utilize the ethanol produced after the glucose in the medium had been exhausted. Though this absence of ethanol utilization might be a good industrial phenotype since it increases the yield of ethanol during fermentation, this might also be associated to the respiratory growth defect, which might compromise large-scale yeast propagation for industrial application.

To assess whether the absence of growth after the diauxic shift is associated with a respiratory growth defect, the strain was first tested for growth in either ethanol or glycerol as a sole carbon source. The assay was done in liquid YP medium containing 5% ethanol or glycerol as a sole carbon source. Accordingly the evolved strain GS1.11-26 was not able to grow in medium containing only ethanol as a carbon source (Figure 3.7 a). On the other hand, it was able to grow in medium containing glycerol as sole carbon source, but with a slower growth rate than the parent strain HDY.GUF5 (Figure 3.7 b). The ability to grow in glycerol indicated that the

strain was not completely defective for respiration, but that its maximal rate of respiration was significantly reduced compared to the parent strain. The complete lack of growth in ethanol might be due to a specific additional inability of the strain to metabolize ethanol as a carbon source, and could then be the reason for the absence of diauxic shift.

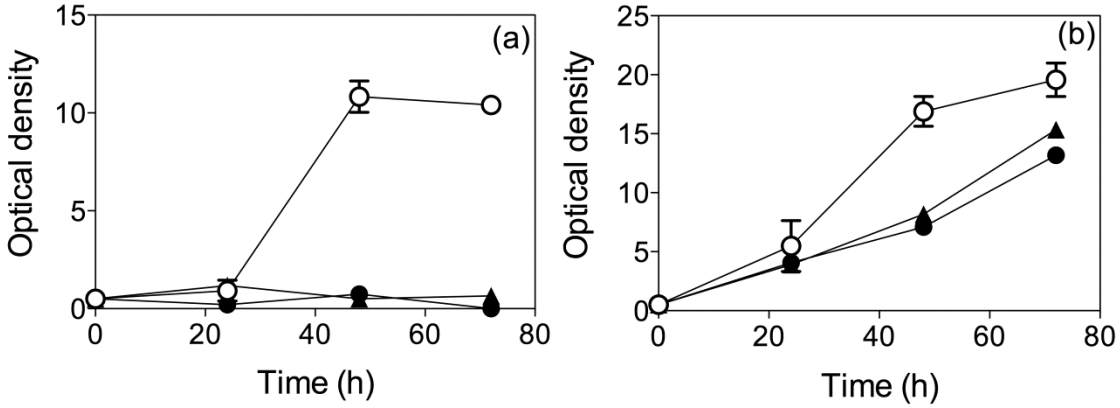


Figure 3.7 . Growth rate of strain GS1.11-26 under aerobic conditions. Growth in shake flask cultures at 30 °C with YP medium containing 5% ethanol (a) or 5% glycerol (b) as sole carbon source. Strains were inoculated at an initial OD of 0.5 and tested in duplicate. The OD was normalized by subtracting growth in YP medium without added carbon source. Error bars represent standard error of the mean from duplicate experiments. Strains: (●) GS1.11-26, (○) parent HDY.GUF5, (▲) M315.

3.3 Discussion

The strain GS1.11-26 has been evaluated for industrially relevant characteristics, which include fermentation performance in three lignocellulosic hydrolysates, aerobic growth rate in glucose, tolerance to high osmotic and ethanol stresses, and ethanol accumulation capacity. GS1.11-26 performed very well in lignocellulose hydrolysates both in SHF and SSF. The yield of ethanol per g consumed sugars, was slightly higher in all the lignocellulose hydrolysates compared to that in synthetic and YP medium. This is probably due to the lower amount of xylitol and glycerol formed, and is consistent with previous results (Karhumaa *et al.,* 2007b; Olofsson *et al.,* 2008b).

In the SHF, it reached high maximum D-xylose consumption rates of 1.1 g/g inoculum DW/h and it showed partial co-fermentation of glucose and D-xylose. We have used a parameter for calculation of the specific sugar consumption rate based on the initial inoculum density, since the whole slurry was used for the fermentation experiment and since it is difficult to estimate the biomass during the fermentation process. The ethanol yield from glucose and D-xylose in lignocellulose hydrolysates was also close to maximum and final ethanol titers between 3.9 and 5.8% (v/v) were reached, depending on the type of hydrolysate.

SSF is an interesting process for production of ethanol from lignocellulosic feedstocks, e.g. because it strongly reduces feedback-inhibition on enzymatic hydrolysis by the liberated monosaccharides and also reduces the risk of contamination. SSF performed at higher temperature (39°C) was also shown to increase the final yield of ethanol, because of a better compromise between the temperature optima of the enzymes and the yeast (Mutturi and Lidén, 2013). In this respect, GS1.11-26 performed very well with almost complete attenuation of the free glucose and D-xylose in about 96 h at 39°C. In a previously reported SSF of Arundo hydrolysate using Ethanol Red (Mutturi and Lidén, 2013), there was also some D-xylose consumption, but only as a result of D-xylose reduction to xylitol. In our study, the final ethanol concentration for GS1.11-26 was 20.3 g/L, corresponding to an ethanol yield of 0.29, to be compared to the previously reported values for Ethanol Red of 15.3 g/L and 0.22 g ethanol/g total sugar. The ethanol yield thus increased by about 32% due to the efficient D-xylose conversion. The increase in the ethanol yield was in fact slightly higher than the increase expected from the D-xylose conversion alone, possibly due to removal of D-xylose inhibition on enzymatic hydrolysis.

The GS1.11-26 strain maintained a high level of tolerance like its parent Ethanol Red in inhibitor rich spruce hydrolysate and to individual inhibitors HMF and furfural. However, the strain did not retain the same high ethanol tolerance as the original Ethanol Red parent strain, though it was still able to accumulate more than 15% ethanol in very-high gravity fermentation (YP + 330 g/L glucose). The relatively high tolerance of GS1.11-26 to inhibitors, like HMF and furfural,

found in high concentration in spruce hydrolysate, but its lower tolerance to other stresses, like ethanol and acetic acid, can be explained by the fact that, after the genome shuffling step spruce hydrolysate was used as selective medium. The cells that were able to grow in spruce hydrolysate were further used for the evolutionary adaptation. This result demonstrates the importance of the selection conditions during evolutionary engineering, which is in agreement with the principle, "you get what you screened for" (Schmidt-Dannert and Arnold, 1999). In addition, GS1.11-26 showed reduced tolerance to acetic acid compared to the parent HDY.GUF5. This can be explained by different mechanisms underlying tolerance to various inhibitors (Almario *et al.,* 2013b). In *S. cerevisiae*, the tolerance mechanism to HMF and furfural, is similar, but distinct from that of acetic acid (Palmqvist and Hahn-Hägerdal, 2000).

Overall, GS1.11-26 demonstrated an efficient fermentation rate of glucose and D-xylose in the three lignocellulose hydrolysates tested. However, it showed a reduced aerobic growth rate and a slightly reduced glucose fermentation rate compared to the parent strain HDY.GUF5. Moreover, the D-xylose utilization capacity especially in SSF of spruce hydrolysate has been reduced in comparison to the rate in complex medium, probably due to the reduction in the inhibitor tolerance, especially to acetate. In spite of this, GS1.11-26 has a significant potential for further development of a robust industrial yeast strain for bioethanol production with various lignocellulose hydrolysates.

Chapter 4

Development of an industrial *S. cerevisiae* strain that combines efficient D-xylose utilization capacity with inhibitor tolerance and other industrially important traits

Adapted from:

Demeke MM, Dumortier F, Li Y, Broeckx T, Foulquié-Moreno M and Thevelein JM. **"Combining inhibitor tolerance and Dxylose fermentation in industrial Saccharomyces cerevisiae strain for efficient lignocellulose based bioethanol production**". *Biotechnology for Biofuel*, 2013, **6**:120

Author's contribution

The author participated in the design of the project, the experimental work, data analysis and drafted the manuscript.

3.1 Introduction

The main challenges in advanced bioethanol production are the development of efficient and cheap technologies to liberate all fermentable sugars from lignocellulosic feedstocks, and the engineering of robust microorganisms able to rapidly ferment all sugar present in the biomass hydrolysate, mainly glucose and D-xylose (Schubert, 2006; Weber *et al.,* 2010). Several pretreatment and enzymatic hydrolysis processes have been reported with increasing efficiency for releasing the sugars from the biomass (Galbe and Zacchi, 2007; Jørgensen *et al.,* 2007; Taherzadeh and Karimi, 2008). Yeast strains have even been developed that secrete cellulolytic enzymes for use in consolidated bioprocessing (van Zyl *et al.,* 2007). However, in addition to the release of fermentable sugars, large amounts of several types of inhibitory compounds are released during the pretreatment process. They inhibit microbial fermentation and growth, resulting in severely reduced ethanol yield and productivity. Therefore, economically viable industrial production of lignocellulose-derived bioethanol requires not only a microorganism that is able to ferment all hexose and pentose monosaccharides in the lignocellulose hydrolysates, but also exhibits unusually high tolerance to the toxic compounds present in the lignocellulose hydrolysates.

Substantial progress has been made in the past few years to develop yeast strains that are able to ferment D-xylose(Bettiga *et al.,* 2008; Brat *et al.,* 2009; Hahn-Hägerdal *et al.,* 2007a; Kuyper *et al.,* 2005a; Wahlbom *et al.,* 2003; Zhou *et al.,* 2012), and to obtain strains with improved inhibitor tolerance (Gorsich *et al.,* 2006; Larsson *et al.,* 2001b; Larsson *et al.,* 2001a). Some D-xylose fermenting recombinant strains of *S. cerevisiae* and natural D-xylose utilizing yeast species with improved inhibitor tolerance have also been reported (Hou and Yao, 2012; Koppram *et al.,* 2012). However, most of these accomplishments have been obtained with laboratory *S. cerevisiae* strains, or strains of *S. cerevisiae* and other yeast species without proven track record in industrial bioethanol production. In addition, the performance of the best strains available in terms of D-xylose fermentation and inhibitor tolerance still requires much improvement in order to reach efficient fermentation of lignocellulosic hydrolysates, especially at a higher solid loading (Olofsson *et al.,* 2008a). Since pentose fermentation appears to be much more sensitive to the toxic inhibitors (Bellissimi *et al.,* 2009), the productivity of the yeast in high-density lignocellulose hydrolysates is largely determined by the robustness of the pentose fermentation.

In the previous chapters, the construction and evaluation of the D-xylose fermenting strain GS1.11-26 has been described. This strain has been constructed from Ethanol Red, a prime industrial yeast strain used in first-generation bioethanol production with corn and wheat. For that reason, the strain GS1.11-26 was considered to have very promising potential for development of an all-round robust yeast strain for efficient fermentation of various

lignocellulosic materials. However, due to the accumulation of background mutations during the mutagenesis and/or evolutionary engineering procedures used to develop the strain (chapter 2), GS1.11-26 did not retain the same tolerance to high concentrations of ethanol and acetate, and showed reduced ethanol accumulation capacity in very high-density fermentations compared to the original Ethanol Red background strain HDY.GUF5 (chapter 3). Moreover, it also had a partial respiratory defect causing a reduced aerobic growth rate, which would compromise large-scale propagation of yeast in fed-batch mode. Hence, as such the strain would not be suitable for direct industrial application.

In this chapter, we report the development of three new D-xylose-utilizing industrial yeast strains, derived from the GS1.11-26 strain, which lack its negative properties. The new strains are diploid and were obtained through meiotic recombination with a diploid segregant (derived from a strongly inhibitor-tolerant triploid strain) and with a haploid segregant of Ethanol Red. The three new superior strains exhibited significantly increased tolerance to various inhibitors in spruce hydrolysate, faster growth rate in glucose medium and a faster glucose consumption rate and higher ethanol accumulation capacity in very high gravity fermentations. The maximum D-xylose utilization rate of the three new strains was slower than that of GS1.11-26, but they completely consumed 37 g/L D-xylose and 36 g/L glucose in about 32 h. Our results also demonstrate that commercially important traits present in diploid industrial yeast strains can be combined into a single industrial yeast strain with superior properties and performance without the need for isolation of haploid derivatives.

4.2 Meiotic recombination of GS1.11-26 with a superior inhibitor tolerant strain

4.2.1 Screening of an *S. cerevisiae* strain collection for tolerance to inhibitors in spruce hydrolysate

We first aimed at obtaining a strain with extremely high performance in terms of growth and fermentation directly in inhibitor-rich lignocellulose hydrolysate, since simultaneous tolerance to multiple inhibitors is important for high productivity in lignocellulose hydrolysates (Zheng *et al.,* 2011). For that purpose, we chose a robust inhibitor tolerant strain (JT21653) which has previously been selected by our research group (in a project led by Dr. Françoise Dumortier) from the screening of 580 yeast strains for inhibitor tolerance in undetoxified acid pretreated spruce hydrolysate.

JT21653 is a baker's yeast that was purchased from a local commercial source. A species identification test was performed on a single cell isolate at the BCCM/MUCL (Mycothèque de l'Université Catholique de Louvain, Louvain-la-Neuve, Belgium), which confirmed that the yeast was *S. cerevisiae* Meyen ex. E.C. Hansen. Flow cytometry analysis of the DNA content revealed

that JT21653 had a triploid genome (Figure 4.1). The strain showed good sporulation and high spore viability and hence, was selected for further analysis.

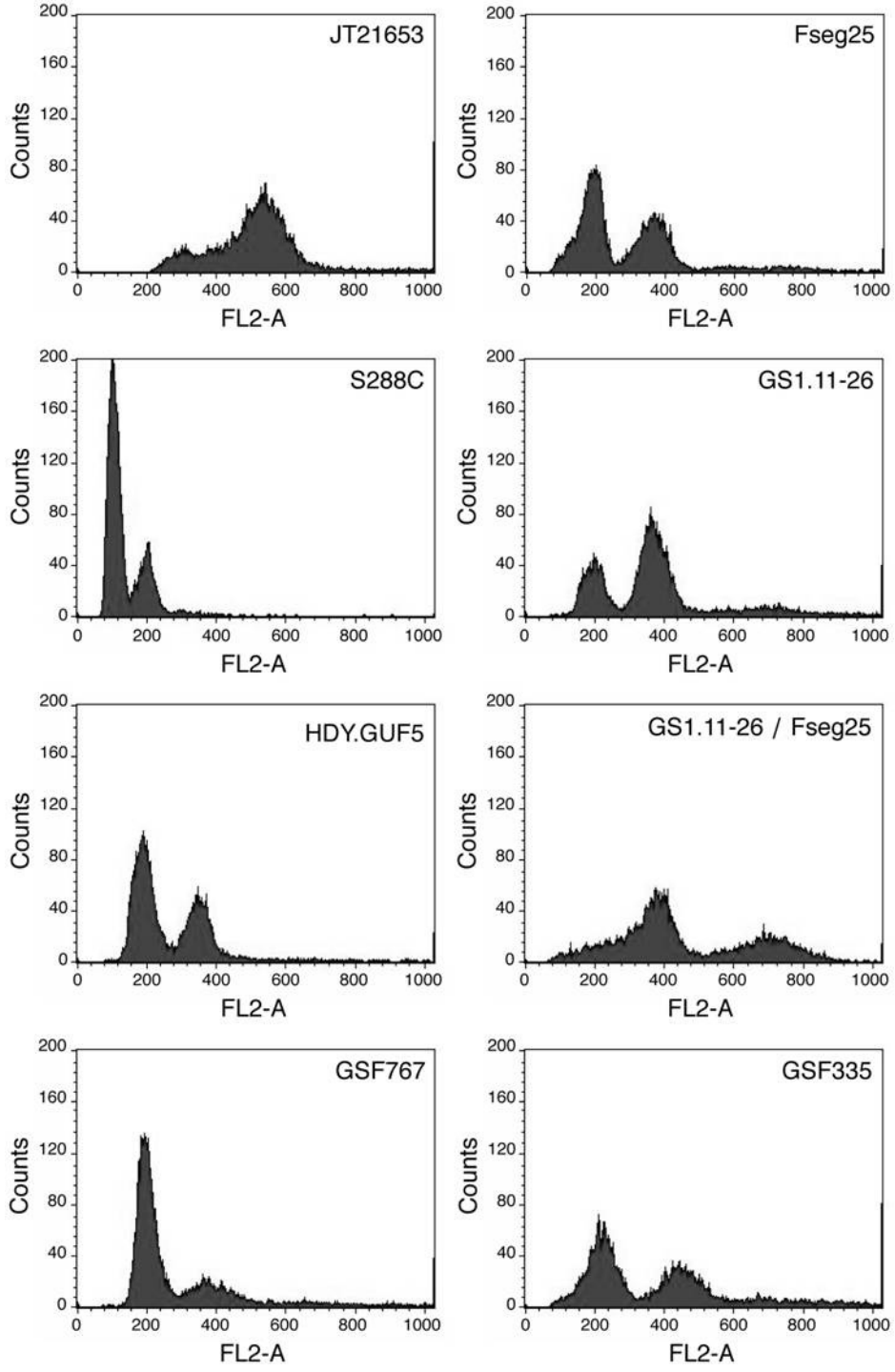


Figure 4.1. Determination of DNA content by flow cytometry. Strains were grown to exponential phase, after which they were fixed with ethanol, and the DNA was stained with propidium iodide. Control haploid (S288c) and diploid (HDY.GUF5) strains were used for comparison.

4.2.2 Isolation of an inhibitor tolerant segregant of JT21653

In order to select a mating competent segregant of JT21653 with an inhibitor tolerance at least as good as the triploid parent, we screened 41 segregants of JT21653 for fermentation performance under semi-anaerobic conditions in 40% spruce hydrolysate supplemented with YP and glucose to 200 g/L at pH 5.5 (results not shown). From this prescreening, we selected the 15 segregants with the best fermentation performance. For selection of the most inhibitor-tolerant segregant, the concentration of the spruce hydrolysate was increased to 77% and supplemented with 70 g/L glucose. This concentration of spruce hydrolysate was severely inhibitory to the industrial strain Ethanol Red, the triploid parent JT21653 as well as the majority of the 15 segregants, whereas four segregants were able to complete the fermentation under these conditions albeit with varying rates (Figure 4.2). One segregant, named Fseg25 (which has mating type *MAT**a***), performed exceptionally well, initiating the fermentation with virtually no lag phase and completing it already after about 48h, a time at which only one other segregant had made a significant start-up of the fermentation. Analysis of the DNA content of Fseg25 by flow cytometry indicated that the strain had a diploid genome (Figure 4.1). In addition, it was able to mate with a *MATα/α* strain, indicating that it was *MAT**a/a.*** The resulting tetraploid strain was able to produce viable spores. Therefore, the Fseg25 segregant was chosen as a mating partner for genetic recombination with the diploid *MATα/α* strain GS1.11-26.

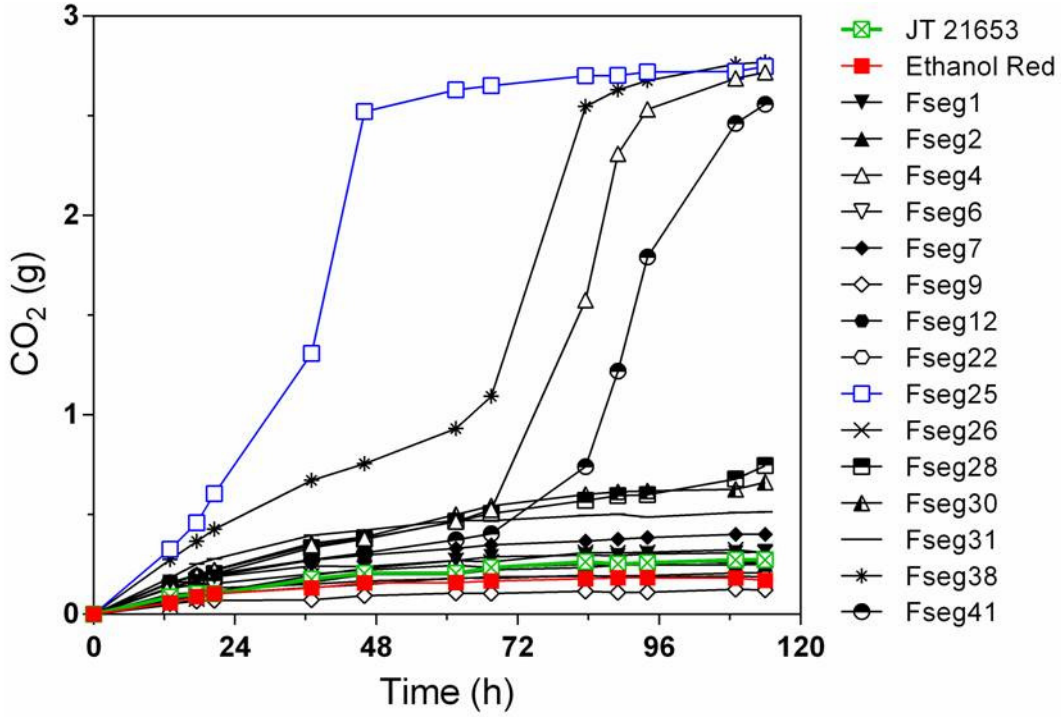


Figure 4.2. Selection of a superior segregant from the inhibitor tolerant baker's yeast JT21653. Fermentation was performed in semi-anaerobic conditions using acid-pretreated slurry of spruce material at 77% concentration and supplemented with 70 g/L glucose. CO_2 production was estimated from the weight loss. Fseg25 (blue) showed the best performance, with in particular a very short lag phase compared to the other segregants. The diploid parent

JT21653 (green) and the industrial bioethanol production strain Ethanol Red (red) as well as all other segregants, except for four, were not able to ferment appreciably at this concentration of spruce hydrolysate within 120 h.

4.2.3 Meiotic recombination of GS1.11-26 with Fseg25

To combine the superior D-xylose fermentation performance of GS1.11-26 with the superior inhibitor tolerance trait of Fseg25, we mated the *MATα/α* strain GS1.11-26 with the *MAT**a**/**α*** segregant Fseg25. The tetraploid strain generated was sporulated, and 819 meiotic segregants were screened in order to select diploid hybrid F1 segregants with the required properties, mainly efficient D-xylose fermentation, fast aerobic growth and high inhibitor tolerance in spruce hydrolysate.

4.2.4 Screening for D-xylose fermentation capacity

We first performed a prescreening of the 819 segregants based on their ability to grow on YPX solid medium, in order to reduce the number of strains to be evaluated in fermentations. All segregants that showed detectable growth on solid D-xylose medium were further analyzed for growth in liquid medium. For that purpose, the strains were inoculated in 1 ml YPX medium at an initial OD_{600} of 1.0. After about 24 h of incubation, a range of cell densities, from OD_{600} of about 5 up to 33 was observed for the different segregants. Strain GS1.11-26 showed an OD_{600} between 28 and 33 in different replicate growth assays. To monitor the correlation between growth in liquid YPX and fermentation performance, segregants growing to an OD_{600} above 5 were evaluated by fermentation in YP medium containing 40 g/L D-xylose. We observed that, most of the best D-xylose fermenting strains also performed well in such growth evaluation experiments (data not shown). Thus, the majority of poor D-xylose fermenting segregants could be excluded by using a cut-off value for growth to an OD_{600} of 15 in 24 h, since all the good D-xylose fermenting segregants grew to an OD_{600} of above 15. Hence, growth in liquid YPX medium for 24h and selection of the segregants growing to a minimum OD_{600} of 15 was considered to be the best method for rapid initial screening and elimination of poor performers.

Using this method, about 168 segregants growing to OD_{600} values of about 15 in 24 h were preselected and further tested for D-xylose fermentation performance in semi-anaerobic conditions. This was done in different batches of experiments (results not shown) and finally resulted in 48 segregants with moderate to rapid D-xylose fermentation capacity (Figure 4.3). To allow a proper comparison, the 48 selected segregants were evaluated in a single batch of fermentation experiments. The 27 best segregants, with a D-xylose fermentation performance close to that of GS1.11-26, were eventually selected for further analysis (Figure 4.3). Flow cytometry analysis showed that the selected 27 segregants all had a DNA content similar to that

of a diploid control strain (data not shown). Hence, all segregants appeared to be diploid strains, although aneuploidy for one or more chromosomes cannot be ruled out.

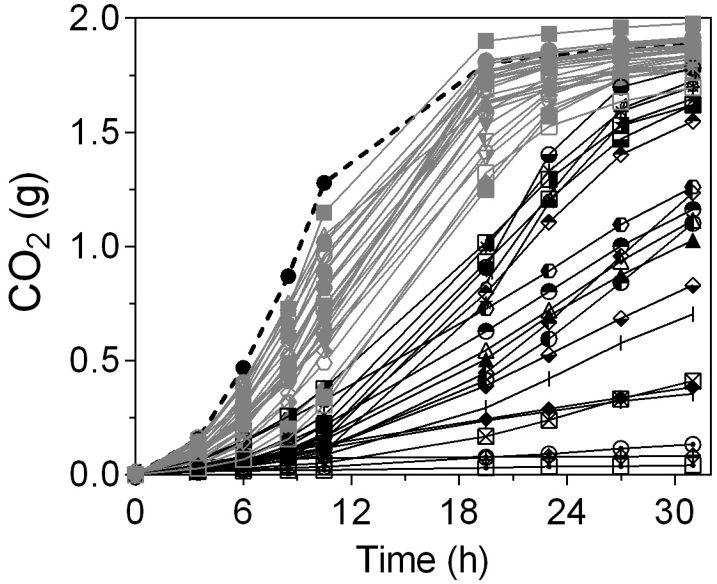


Figure 4.3. D-xylose fermentation profile of the 48 preselected D-xylose utilizing segregants obtained from the tetraploid strain GS1.11-26/Fseg25, in YP medium containing 40 g/L D-xylose. The 27 selected segregants with the best performance are shown in grey lines and the parent strain GS1.11-26 is shown broken lines. A similar fermentation profile was obtained during the preselection step in the same conditions.

4.2.5 Screening of hybrid segregants for aerobic growth rate and inhibitor tolerance

One of the shortcomings of GS1.11-26 for direct industrial application was the slow aerobic growth rate, which is a key factor for yeast propagation in the preparation of sufficient inoculum to start the industrial fermentation process. Hence, we screened the 27 D-xylose fermenting strains for growth rate in synthetic medium containing glucose as a carbon source. The recombinant strain HDY.GUF5 (Ethanol Red background) was used as a reference throughout this chapter since GS1.11-26 was developed from this strain background (chapter 2). From the 27 segregants, only six showed a growth rate as high as that of the HDY.GUF5 strain (Figure 4.4). When these six segregants were evaluated in inhibitor-rich spruce hydrolysate, only two segregants, named GSF335 and GSF767, grew better than the HDY.GUF5 strain (data not shown). Both strains were found to be diploid and had MATα/α mating type. They were selected for further evaluation.

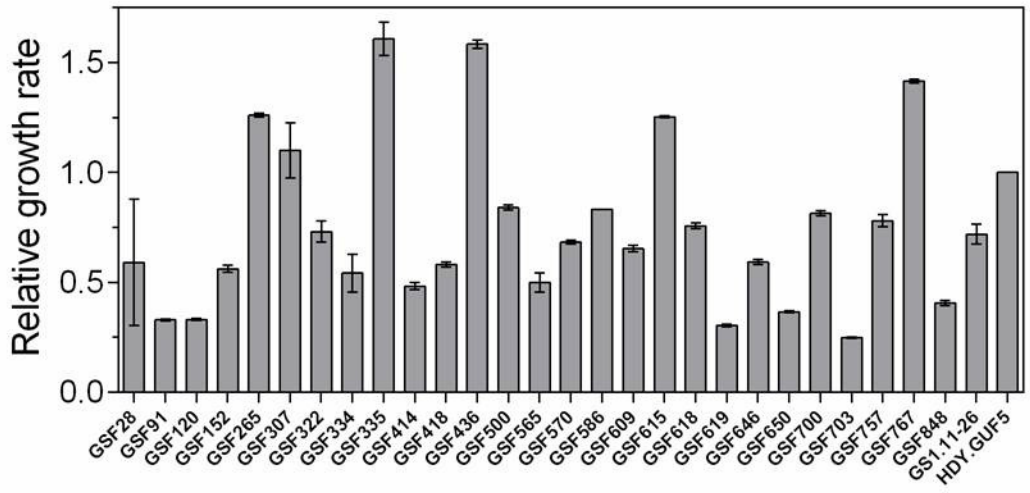


Figure 4.4. Relative maximum growth rate of the 27 best D-xylose fermenting segregants in comparison to the strain HDY.GUF5. Growth was performed in synthetic medium containing 20 g/L glucose in 200 µl volume using a bioscreen assay at an initial OD_{600} of 0.1. Error bars represent standard error of the mean from average values of triplicate experiments. The maximum growth rate obtained for each strain was calculated relative to that of HDY.GUF5.

4.3 Backcrossing the diploid GS1.11-26 with a haploid segregant of Ethanol Red

In a parallel approach, the MATα/α diploid strain, GS1.11-26, with high D-xylose fermentation activity, was crossed with the segregant ER17 of Ethanol Red. ER17 was previously selected among a number of segregants for its higher acetic acid tolerance in fermentation, which was even better than that of its diploid parent strain Ethanol Red (Meijnen et al, unpublished data).

The crossing of GS1.11-26 with ER17 resulted in a triploid strain. Meiotic segregation in triploid strains often produces two haploid and two diploid progeny in a tetrad (Charles *et al.,* 2010). Obtaining diploid progeny is important since haploid strains are unattractive for industrial application, because of their lower genetic stability and robustness (Ding *et al.,* 2010). The triploid strain was sporulated and 104 segregants were isolated. Most of the tetrads produced four viable spores. All the isolates were then evaluated for the important phenotypic traits.

4.3.1 Screening of the hybrid segregants from GS1.11-26/ER17 for growth in D-xylose medium

The 104 segregants isolated from GS1.11-26/ER17 were first screened for growth in YPX medium as a preselection step. A range of final OD_{600} values from about 2 to 33 was observed (Figure 4.5), indicating the involvement of multiple genetic factors for D-xylose growth. Since

growth on and fermentation of D-xylose have previously shown a good correlation, the 21 segregants that grew best in D-xylose medium were selected for the next evaluation.

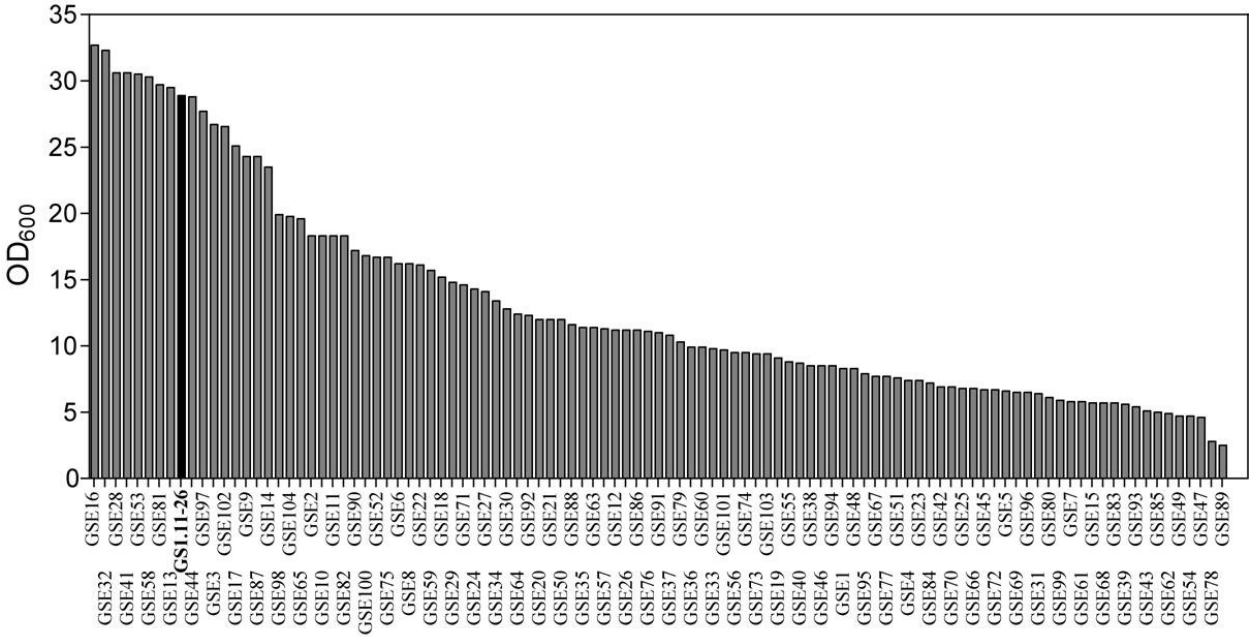


Figure 4.5. Screening of 104 segregants obtained from the triploid strain GS1.11-26/ER17, for growth in 1ml YP medium containing 20 g/L D-xylose as carbon source. Cells were inoculated at an OD_{600} value of 1.0. The final OD_{600} was then measured after 20 h. The parent strain GS1.11-26 was used as a reference.

4.3.2 Screening of the hybrid segregants from GS1.11-26/ER17 for aerobic growth rate

The 21 best D-xylose growing segregants were further tested for growth rate in glucose medium. They were pregrown in 3 ml YPD medium. Segregants that yielded after 16 h much lower biomass than HDY.GUF5 were excluded. The remaining seven segregants were tested for growth rate in synthetic medium containing 20 g/L glucose. Two segregants that produced after 16 h lower OD_{600} values than HDY.GUF5 were included for comparison.

In the growth rate assay, seven of the nine strains tested showed a much faster growth rate than GS1.11-26 (Figure 4.6). The two segregants, GSE17 and GSE102, that produced lower OD_{600} values during the preselection step, also showed a slower growth rate than GS1.11-26, confirming the results from the pregrowth assay. The seven segregants that grew with a rate close to that of HDY.GUF5 were then selected for the final fermentation experiment in D-xylose medium.

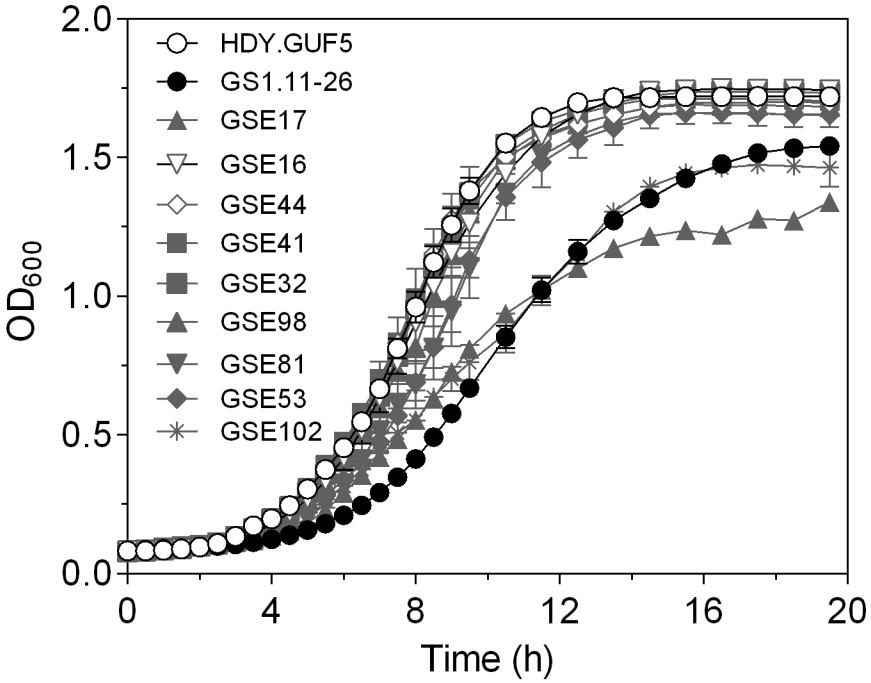


Figure 4.6. Screening of the best D-xylose growing segregants, obtained from the triploid strain GS1.11-26/ER17, for aerobic growth rate in glucose medium. A bioscreen assay was performed in 200µl volume synthetic medium containing 20 g/L glucose, at an initial OD_{600} of 0.1. Error bars represent standard deviation from the average of triplicate experiments.

4.3.3 Selection of the most superior D-xylose fermenting and inhibitor tolerant hybrid strains

To select first the best D-xylose fermenting strains, the seven hybrid strains with a rapid growth rate in glucose were evaluated for fermentation performance in 50 ml YP medium containing 40 g/L D-xylose as a sole carbon source. The original evolved strain GS1.11-26 was used for comparison. As can be seen in figure 4.7, two strains, GSE44 and GSE16, utilized D-xylose very well, nearly as good as GS1.11-26. Their rate of D-xylose utilization (estimated from the CO_2 production rate) was slightly lower than that of GS1.11-26.

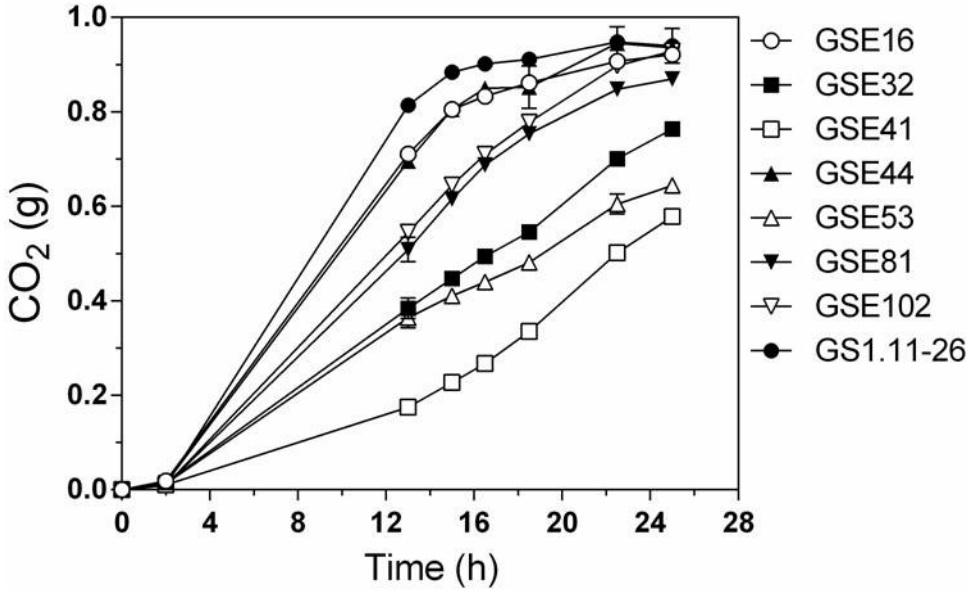


Figure 4.7. Selection of efficient D-xylose fermenting segregants from the seven segregants with best performance, obtained from the triploid strain GS1.11-26/ER17. The fermentations were performed in 50 ml volume with YP medium containing 40 g/L D-xylose. The CO_2 release was estimated from the weight loss during fermentation. Error bars represent standard error of the mean from the average in duplicate experiments.

These two strains, GSE44 and GSE16, were tested for tolerance to acetic acid and to spruce hydrolysate. GSE16 showed similar tolerance towards acetic acid and inhibitors in spruce hydrolysate as the HDY.GUF5 strain (growth in liquid YPD containing 6 g/L acetic acid, pH4.5, or 80% of the liquid portion of spruce hydrolysate, pH 5). On the other hand, GSE44 displayed reduced tolerance especially to acetic acid, showing growth in the presence of only up to 4 g/L acetic acid and no growth anymore at 5 g/L acetic acid (results not shown). Hence, the diploid strain GSE16, which is *MAT**a**/α*, was selected for further evaluation.

4.4 Evaluation of the most superior hybrid strains

4.4.1 Fermentation performance with a glucose-D-xylose mixture

The three selected strains, GSE16 (obtained from backcrossing GS1.11-26 with a segregant of Ethanol Red), GSF335 and GSF767 (obtained from crossing GS1.11-26 with a segregant of JT21653), were evaluated in a more controlled fermentation experiment in YP medium containing 36 g/L glucose and 37 g/L D-xylose at 35 °C. The inoculum cell density was 1.3 g DW/L. The performance of the three hybrid strains was compared with that of GS1.11-26 (which was shown in chapter 2). GS1.11-26 completely consumed both D-xylose and glucose in about 13 h, compared to about 32 h for the other three strains (Figure 4.8).

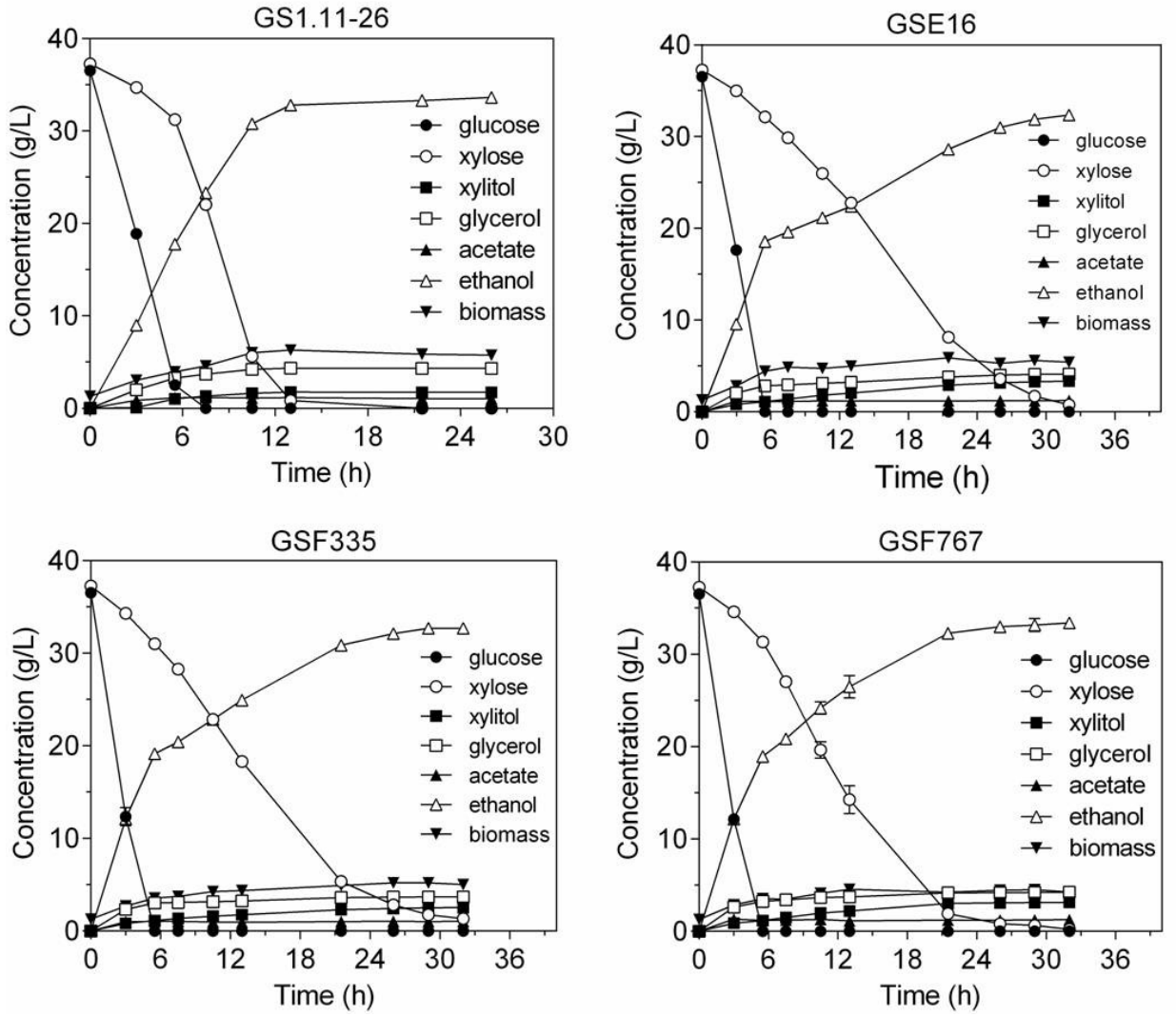


Figure 4.8. Fermentation performance of the three new hybrid strains in a glucose and D-xylose mixture in comparison to that of GS1.11-26. The fermentations were performed in YP medium containing 37 g/L glucose and 37 g/L D-xylose, with an initial biomass of 1.3 g/L. Error bars represent standard error of the mean from the average in duplicate experiments.

The glucose consumption rate with the three new hybrid strains was higher than that of the evolved strain GS1.11-26 (Figure 4.8 and 4.9a). Moreover, GSF335 showed even a slightly higher glucose consumption rate than the original parent strain HDY.GUF5 (Figure 4.9a). However, GS1.11-26 remained superior in terms of D-xylose consumption rate and ethanol productivity from D-xylose (Figure 4.8 and 4.9b). Among the three new hybrid strains, GSF767 showed the highest D-xylose consumption rate (0.65 g/gDW/h), but its maximum D-xylose utilization rate was still about 40% lower than that of GS1.11-26 (1.10 g/gDW/h). On the other hand, the same ethanol yield was obtained with GS1.11-26 and GSF767 (0.46 g/g initial sugar or 90.2% of the theoretical maximum), while it was slightly lower with the other two strains GSF335 and GSE16 (0.44 g/g initial sugar or 86.3% of the theoretical maximum). Partial co-

utilization of D-xylose and glucose was observed with all three new hybrid strains as well as with GS1.11-26 (Figure 4.8).

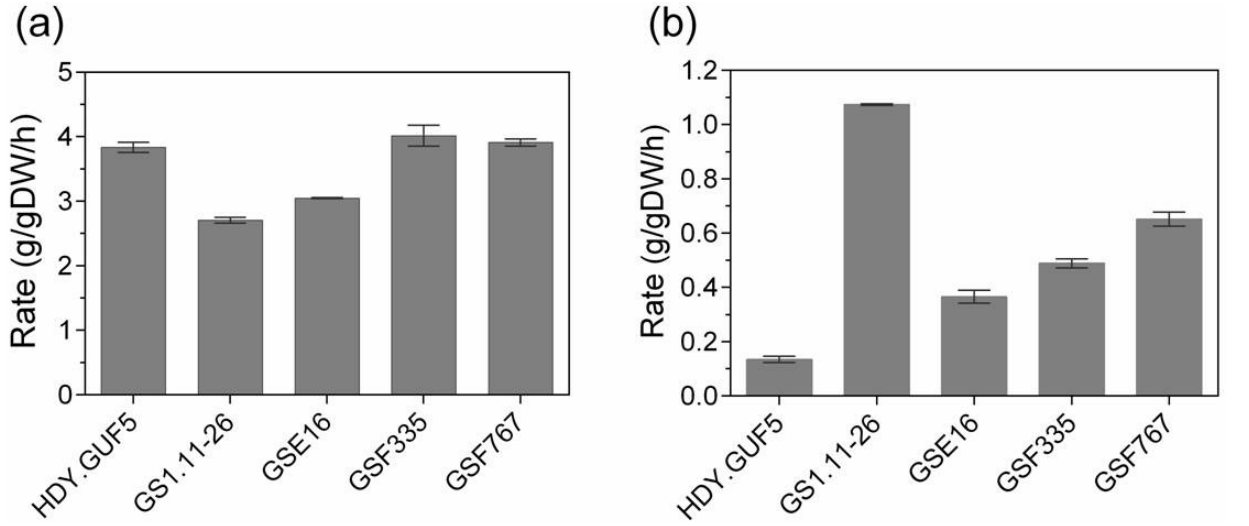


Figure 4.9. Maximum sugar consumption rate attained by the three new hybrid strains and GS1.11-26. (a), maximum glucose consumption rate; (b), maximum D-xylose consumption rate. The values were calculated from the fermentation experiment shown in Figure 9. Maximum D-xylose utilization rate was obtained after glucose exhaustion. Error bars represent standard error of the mean from the average in duplicate experiments.

4.4.2 Fermentation performance in inhibitor-rich spruce hydrolysate

The three new hybrid strains, GSE16, GSF335 and GSF767, were evaluated for inhibitor tolerance in a fermentation experiment with acid pretreated spruce material. The whole slurry of the spruce material, 60%, supplemented with yeast extract and peptone, was used. Glucose (40g /L) was added since the sugar concentration in the hydrolysate was rather low (about 13 g/L), which would not allow a proper comparison of the fermentation performance between the strains. Under these conditions, the rate of fermentation (as estimated from CO_2 release) by GSF335 and GSE16 was much faster than that of GS1.11-26 (Figure 4.10). GSF767 also performed slightly better than GS1.11-26. This result indicates that the new hybrid strains, especially GSF335 and GSE16, have significantly better tolerance than GS1.11-26 towards the inhibitors present in acid-pretreated spruce hydrolysate.

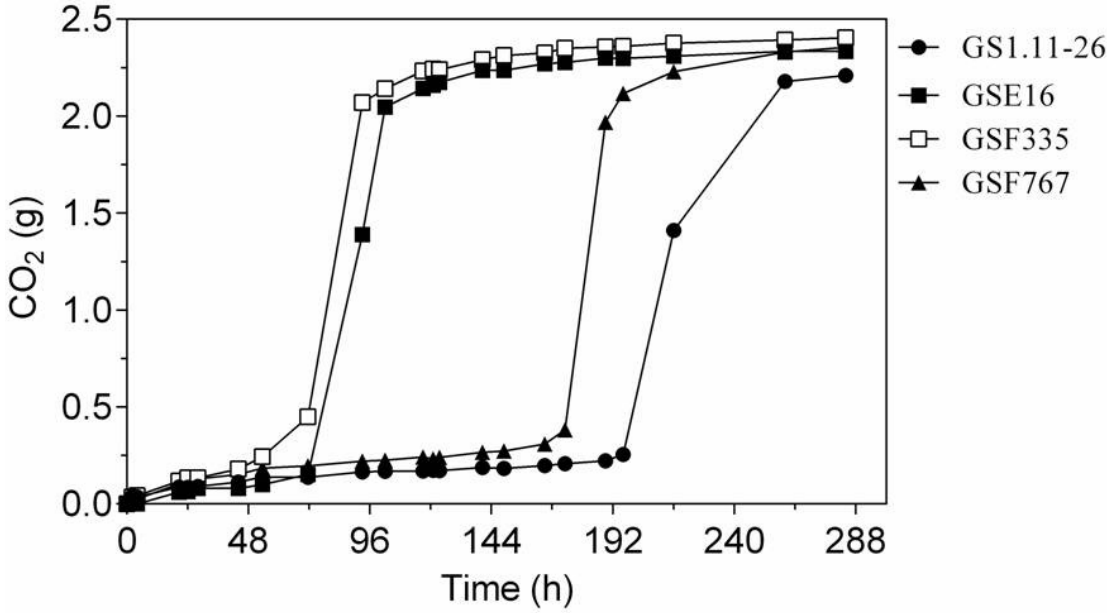


Figure 4.10. Inhibitor tolerance assay in acid-pretreated spruce hydrolysate with the three new hybrid strains in a small-scale fermentation experiment. The fermentation was performed using the whole slurry of spruce hydrolysate (60%) supplemented with 40 g/L glucose with continuous stirring at 200 rpm. Cells were inoculated at an initial cell density of 1.3 g DW/L from a stationary phase pre-culture in YPD medium. The CO_2 production was estimated from the weight loss during fermentation.

To evaluate more precisely the D-xylose and glucose utilization rates in the spruce hydrolysate, we studied the performance of the three strains in a similar batch fermentation using undetoxified spruce hydrolysate at a solid loading of 11% (corresponding to about 50% of the slurry), and supplemented with glucose and D-xylose to a final concentration of 62 g/L and 18 g/L, respectively. The fermentation was performed at an initial cell density of 4 g DW/L (Figure 4.11). Glucose and mannose (derived from the hydrolysate) were completely consumed in less than 10 h by all three strains as well as by the control strains HDY.GUF5 and GS1.11-26, with GSF767 showing the fastest glucose consumption rate, (consumed all the glucose in 4 h). The D-xylose fermentation rate was much slower compared to that of glucose fermentation and was now much more similar for all strains than in the more concentrated hydrolysate (except HDY.GUF5, which cannot use D-xylose). The hybrid strain GSF767 showed a similar D-xylose fermentation rate to that of GS1.11-26, whereas the rate was slightly lower for GSE16 and GSF335. No xylitol and little glycerol (about 0.05 g/g) was produced by the three hybrid strains and GS1.11-26 (results not shown). Although the D-xylose utilization rate was slower, the final ethanol concentration and yield obtained by the hybrid strains and GS1.11-26 was comparable. Compared to HDY.GUF5, the two hybrid strains GSE16 and GSF767 as well as GS1.11-26 produced about 23% more ethanol, due to their efficient D-xylose utilization. The ethanol yield of GSF335 was lower, but still 13% higher than that of HDY.GUF5.

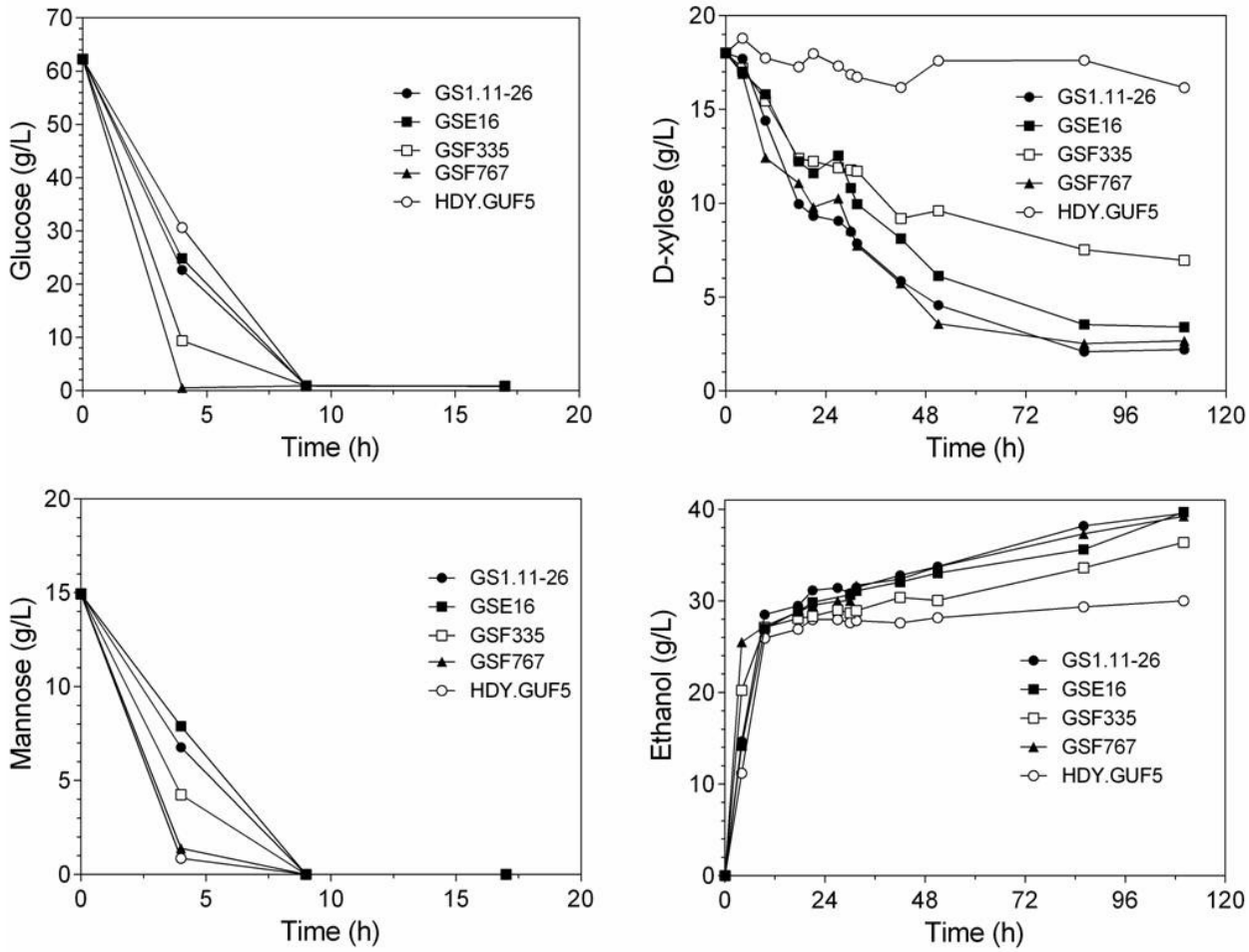


Figure 4.11. Fermentation profile of the three new hybrid strains in semi-anaerobic batch fermentation with acid-pretreated spruce hydrolysate. The fermentation was performed with acid-pretreated spruce hydrolysate at a solid loading of 11%, supplemented with yeast nitrogen base, ammonium sulfate and amino acids. The final glucose and D-xylose concentrations were adjusted to 62 g/L and 18 g/L, respectively. The mannose was only derived from the hydrolysate. Fermentation was started at an initial density of 4 g DW/L and carried out with continuous stirring at 200 rpm.

4.4.3 Fermentation performance in very-high gravity fermentation

The hybrid strains, GSE16, GSF335 and GSF767, were also evaluated for fermentation performance in very-high gravity fermentations using YP medium containing 330 g/L glucose (performed by Yingying Li). The wild type, HDY.GUF5, and the evolved strain, GS1.11-26, were included as controls. The fermentation was performed under conditions of continuous stirring at 120 rpm and under mainly static conditions, with only 4 h of stirring in the beginning. Under both conditions, the three hybrid strains performed much better than GS1.11-26 (Figure 4.12).

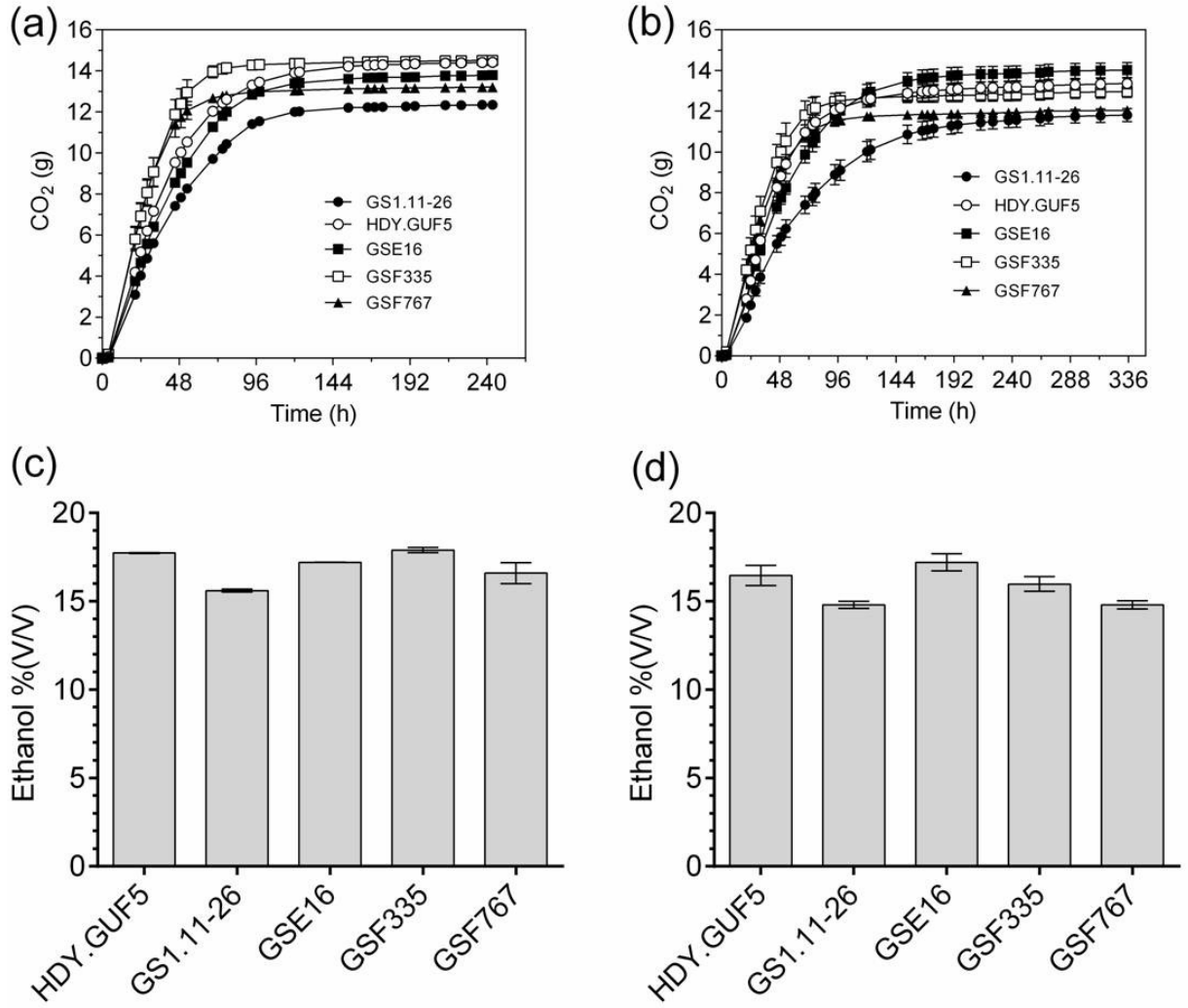


Figure 4.12. Very-high gravity fermentation in YP with 330 g/L glucose. (a) and (b) show the fermentation profile in continuous stirring and static conditions, respectively. The figures in (c) and (d) show the final ethanol titer produced during fermentation of (a) and (b), respectively. Fermentation was started at an initial cell density of 1.3 g DW/L.

Under continuous stirring, GSF335 showed the fastest rate of fermentation (Figure 4.12a). It reached the same final ethanol concentration as HDY.GUF5, but about 96 h earlier. Moreover, GSF335 also showed the highest rate of fermentation under static fermentation conditions. However, the highest final ethanol titer in static conditions was reached by GSE16 (Figure 4.12b and d). GSF767 performed similar to HDY.GUF5 for fermentation rate, but reached a somewhat lower final ethanol titer under both conditions. Overall, the three hybrid strains performed much better than GS1.11-26 under both static and continuous stirring conditions. GSF335 showed superior performance especially in the stirred fermentation conditions accumulating the highest ethanol concentration in only 72 h. In a separate very-high gravity fermentation experiment, the final ethanol titer accumulated by GSF335 was higher than that of both the tetraploid strain GS1.11-26/Fseg25 and the diploid strain Fseg25 (results not shown).

4.5 Discussion

The goal of this study was to develop a superior D-xylose utilizing industrial yeast strain, lacking the negative properties introduced in the background strain GS1.11-26, during the mutagenesis and/or evolutionary engineering procedures used to obtain the strain (chapter 2). The development of the GS1.11-26 strain has shown that introduction of efficient D-xylose fermentation capacity in an industrial yeast strain requires multiple genetic modifications. Hence, just like stress and inhibitor tolerance, efficient D-xylose fermentation is a polygenic trait and the engineering of such traits for development of superior industrial yeast strains has been a major challenge for rational engineering strategies (Attfield, 1997; Borneman *et al.,* 2013; Schwartz *et al.,* 2012; Zhao and Bai, 2009). As a result, most progress in this field has been made up to now with random approaches, like mutagenesis, selection and evolutionary engineering, where the underlying genetic changes responsible for the improvement in performance generally remain unknown (Hou and Yao, 2012; Sonderegger and Sauer, 2003; Wahlbom *et al.,* 2003; Zhou *et al.,* 2012).

In this chapter, we describe the construction of three robust industrial yeast strains that combine efficient D-xylose fermentation with very high inhibitor tolerance. This was achieved by combining a number of strategies. We extended the use of classical breeding with haploid derivatives of the industrial strains to breeding with diploid strains homozygous for the mating locus. This doubles the genetic reservoir for meiotic recombination. In addition, we evaluated large arrays of segregants, using multi-step selection procedures, first aimed at eliminating the truly inferior strains with simple high-throughput tests and gradually moving the selection to conditions as close as possible to the real industrial conditions. In addition, we submitted the candidate strains to very stringent independent evaluation tests, such as the ability to accumulate a very high titer of ethanol in semi-anaerobic very-high gravity fermentations. This is a stringent quality criterion for a superior bioethanol production strain.

We have used meiotic recombination to combine the superior elements from the genomes of the D-xylose utilizing strain GS1.11-26 and the highly inhibitor tolerant strain JT21653. The latter strain was identified through screening of a collection of various yeast strains from a variety of sources using undetoxified acid-pretreated spruce hydrolysate. We chose this material since it has been reported to be among the most inhibitory hydrolysates, containing multiple inhibitors, including the furan derivatives furfural and HMF, aliphatic acids, such as acetic acid, formic acid and levulinic acid, and various phenolic compounds, such as vanillic acid, vanillin, synringaldehyde, syringic acid, and 4-hydroxybenzoic acid (Almeida *et al.,* 2007; Koppram *et al.,* 2012; Persson *et al.,* 2002). A strain tolerant to this medium would most likely also be tolerant to inhibitors in other lignocellulosic hydrolysates that are pretreated in a less severe way. Besides, acid-based thermochemical pretreatment is considered to be one of the

most cost-competitive methods of pretreatment for bioconversion of lignocellulosic feedstocks in liquid biofuel production (Agbor *et al.,* 2011; Galbe and Zacchi, 2007). Hence, such strains would be expected to have broad application potential.

Hybridization of industrial strains is among the most effective and simple techniques used for improving and combining various industrially relevant traits (Pretorius and Bauer, 2002; Yamada *et al.,* 2010). In a classical breeding strategy, haploid strains of opposite mating type are crossed to produce new diploid progeny. However, this method cannot be directly applied to industrial strains since most industrial strains are diploid, polyploid or aneuploid. Hence, the identification of a haploid meiotic segregant with, in principle, the same superior profile of traits as the parent industrial strain is required and this becomes exceedingly difficult when it concerns traits that are only important at industrial scale. Thus, the haploid progeny used for breeding will not display exactly the same repertoire of positive traits as the original diploid strain. This makes breeding of industrial yeast strains a huge challenge. In our work, we have used for mating the original diploid MATα/α strain, GS1.11-26, rather than a haploid derivative, so as to maintain the entire genetic basis underlying the superior D-xylose fermentation capacity of that strain. Constructing hybrid strains by mating two mating-competent diploid parents has been described previously (Hashimoto *et al.,* 2006). However, in that report, no further meiotic recombination step had been applied to the tetraploid hybrid strains to isolate diploid segregants, which might be advantageous for two reasons. First, the most stable genome size in *S. cerevisiae* appears to be the diploid state. This is suggested by the frequent spontaneous evolution of ploidy from tetraploid to diploid in both optimal and stressful environments (Gerstein *et al.,* 2006). This implies that a tetraploid industrial strain will likely show lower stability and might easily lose important traits when losing chromosomes in the shift to a lower ploidy during its industrial usage. Second, meiotic recombination can generate great diversity in a cell population and sporulation of a tetraploid can therefore generate strains with superior performance compared to the tetraploid parent strain (Teunissen *et al.,* 2002). This was also observed in our work, in which diploid segregants were obtained with a much higher D-xylose utilization rate than the tetraploid parent. This is probably due to the involvement of recessive alleles in sustaining efficient D-xylose utilization.

We have used two approaches to improve the GS1.11-26 strain. The backcrossing with an Ethanol Red segregant was mainly aimed at eliminating the negative traits that had been introduced in the background of the strain during its development for high xylose fermentation capacity and inhibitor tolerance, since we maintained the original Ethanol Red background in the breeding process. This approach has been previously described as beneficial for weeding out deleterious mutations in mutant strains (Marullo *et al.,* 2009; Patnaik, 2008). On the other hand, breeding with the Fseg25 strain introduced a new, albeit also industrial, genetic

background and was mainly aimed at further enhancing inhibitor tolerance, with the risk of losing traits important for a bioethanol production strain. Both approaches were successful in generating strongly improved strains. The GSE16 strain was not only cured for the low glucose utilization and aerobic growth rate of its parent GS1.11-26, but also displayed inhibitor tolerance to a level at least as high as the original unevolved parent strain, HDY.GUF5. The GSF335 and GSF767 strains also largely lost the negative properties of their GS1.11-26 parent, but in addition, gained at least part of the very high inhibitor tolerance of the Fseg25 parent strain.

The three hybrid strains were evaluated for various industrially relevant traits, including tolerance to different inhibitors, fermentation performance in a D-xylose/glucose mixture and in very-high gravity fermentation. All three new hybrid strains showed much better general performance in various stress conditions compared to GS1.11-26. There was no single strain that stood out in all conditions. The relative performance varied with the test. In a semi-anaerobic batch fermentation of spruce hydrolysate, the three strains showed a much shorter lag phase compared to GS1.11-26. On the other hand, the very high D-xylose fermentation capacity of GS1.11-26 was not fully maintained in the new hybrid strains. Even though the three hybrid strains were able to consume all the 37 g/L D-xylose and 36 g/L glucose in 32 h, they retained only 35% to 60% of the maximum D-xylose utilization rate of GS1.11-26. The higher inhibitor tolerance of the three new hybrid strains can explain why they displayed D-xylose utilization rates in spruce hydrolysate comparable to that of the strain GS1.11-26. The rate of glucose utilization was also significantly higher than that of GS1.11-26, which is likely also due to their higher tolerance to the inhibitors in spruce hydrolysate. Tolerance to multiple stress factors has previously been shown to correlate with high ethanol yield and a high ethanol production rate in *S. cerevisiae* (Zheng *et al.,* 2011). Moreover, D-xylose fermentation is more sensitive to stress factors, especially to acetic acid, compared to glucose fermentation (Bellissimi *et al.,* 2009). The severe reduction of tolerance to acetic acid (and possibly to other inhibitors) of GS1.11-26 (Chapter 3) can explain its lower D-xylose utilization rate in spruce hydrolysate compared to synthetic medium. On the other hand, the rate of D-xylose utilization by the three new hybrid strains was not as severely reduced as for GS1.11-26 in the spruce hydrolysate compared to complex medium, which can be explained by their higher inhibitor tolerance.

Other diploid hybrid segregants with a higher D-xylose fermentation capacity than the three selected strains have been isolated, but they displayed very slow growth rates in glucose and also reduced inhibitor tolerance. Hence, they were excluded because of these negative properties. The difficulty in maintaining the superior D-xylose fermentation rate of the parent strain GS1.11-26 might suggest that one or more of the negative background mutations in the

strain GS1.11-26 might either be causally or structurally linked to the high D-xylose fermentation rate. If these traits are causally linked, the beneficial genes or loci important for efficient D-xylose fermentation might be linked with the reduced growth rate or with the higher inhibitor tolerance. This would mean that it will not be possible or be very difficult to combine high general robustness and high D-xylose utilization capacity in this strain background. However, if the negative mutations are only structurally linked to the positive genetic modifications, i.e. residing close to each other in the genome, they could be removed without affecting the superior D-xylose fermentation performance. Further research, therefore, should focus on the identification of the genetic basis of efficient D-xylose fermentation as well as the high inhibitor tolerance, so that they can be engineered by reverse metabolic engineering.

Bioethanol production is commonly performed in very-high gravity fermentations using highly concentrated substrates so that a maximal final ethanol titer can be reached (Balat and Balat, 2009). These substrates include first-generation feedstocks, such as sugar cane, starch or grains, and in the future, second-generation feedstocks, in particular lignocellulose waste streams and bioenergy crops. A high final ethanol titer has multiple advantages. It reduces the ethanol distillation costs but also lowers the liquid volumes in the plant causing large savings in heating, cooling, pumping and transport costs. Very-high gravity fermentation causes high stress, in particular osmostress in the beginning of the fermentation and ethanol stress at the end of the fermentation. This results in a longer fermentation time and lower ethanol yield as a result of higher residual sugar (Bai *et al.,* 2008). Hence, the performance of a new yeast strain in very-high gravity fermentation is a crucial quality criterion for its use in real industrial practice. This was especially important since the parent strain, GS1.11-26, was severely compromised in maximal ethanol accumulation capacity in very-high gravity fermentations (chapter 3). Hence, we evaluated the three new hybrid strains also in very-high gravity fermentation and found that they were not only much better than GS1.11-26, but showed even improved perfomance in some cases compared to the original HDY.GUF5 strain. The hybrid strain GSF335 exhibited significantly increased performance compared to its tetraploid parent as well as to both its diploid progenitor strains GS1.11-26 and Fseg25. This clearly shows that meiotic recombination with diploid strains is able to further improve important performance traits that are considered to be already very high in the very best currently used bioethanol production strains.

In conclusion, we have successfully developed three robust industrial yeast strains that combine efficient D-xylose utilization with high inhibitor tolerance for use in bioethanol production with lignocellulose hydrolysates. Two of the strains (GSF335 and GSF767) have been derived through meiotic recombination of the efficient D-xylose utilizing strain GS1.11-26 (derived from Ethanol Red) with the most inhibitor tolerant strain obtained from a screening of more than 580 yeast strains. Strain GSE16 has a purely Ethanol Red background, developed by

backcrossing GS1.11-26 with a haploid derivative of Ethanol Red. All three strains showed superior performance with respect to aerobic growth rate, glucose consumption rate and inhibitor tolerance compared to GS1.11-26. The D-xylose utilization rate of the three strains in complex medium was reduced compared to GS1.11-26, but in inhibitor-rich acid-pretreated spruce hydrolysate it was comparable. Due to the high robustness, the proven record in industrial application of the background strain Ethanol Red, and the efficient D-xylose utilization capacity, the three strains have strong potential for direct application in industrial bioethanol production. This study demonstrates that strains with an optimal profile of industrially important traits can be obtained through meiotic recombination directly with diploid strains and by screening large numbers of strains or segregants in a multi-step selection process, using simple high-throughput screens to eliminate poor performers to more elaborate evaluation conditions, mimicking closely the industrial conditions to select the very best performers.

Chapter 5

Genome sequence analysis of the efficient D-xylose utilizing strain GS1.11-26 to unravel the genetic basis behind its efficient D-xylose utilization capacity

Bibliographic reference

Demeke MM, Foulquié-Moreno M, Duitama J.Dumortier F, Thevelein JM. **Genome analysis of the D-xylose fermenting industrial yeast strain GS1.11-26**. (In preparation)

Author's contribution

The author participated in the design of the project, the experimental work, data analysis and drafted the manuscript.

5.1 Introduction

Because of the complexity and limited understanding about the biochemical and regulatory networks of cellular metabolism, evolutionary engineering has been frequently used as an indispensible method for development of optimized industrial production strains (Petri and Schmidt-Dannert, 2004). The main drawback of such random strain improvement method is the possible side effect of the random genetic modifications, causing undesired phenotypic effects. Moreover, it impairs the rapid identification of the genetic changes responsible for the improved trait, as the modifications are random and spread over the entire genome. We have shown in chapter 3 that the best D-xylose fermenting strain GS1.11-26 exhibited reduced specific growth rate in glucose medium and loss of tolerance towards osmotic and/or ethanol stress in very-high gravity fermentation in medium with 350 g/L glucose. Those traits are industrially important and are required for a strain to be useful in industrial application (Hahn-Hägerdal *et al.,* 2007a). Those negative traits were successfully cured by meiotic recombination with either a robust inhibitor tolerant strain or with a segregant of Ethanol Red (chapter 4). However, the efficient D-xylose utilization capacity of GS1.11-26 was not fully maintained in the new hybrid strains that were cured for the negative traits of GS1.11-26 (chapter 4). Hence, identification of the genetic changes responsible for improved D-xylose fermentation and subsequent transfer of the genetic information into the original industrial strain Ethanol Red can be a good alternative approach to eliminate the undesired phenotypes associated with the random mutations. In addition, the information obtained can be used for improvement of various other industrial production strains. Such a method of constructing a strain with a new trait by random modification, identification of the genetic changes responsible for the trait and endowing the phenotype to a new strain by introducing the identified genetic bases is known as inverse metabolic engineering (Bailey *et al.,* 2002).

The identification of the genetic determinants underlying a specific phenotype is the biggest challenge for inverse metabolic engineering (Nevoigt, 2008). However, the advent of new high-throughput DNA sequencing technologies at a cheap price has facilitated the rapid identification of genetic differences among strains. Global genome sequence comparison was previously shown to be effective in pinpointing major genetic differences between a parent and an evolved strain of *S. cerevisiae* (Araya *et al.,* 2010). We therefore sequenced the genome of the evolved strain and the original parent strain to reveal the genetic differences between them that might be responsible for the high D-xylose fermentation capacity.

Using whole genome sequence comparison between the evolved and the original parent strain, amplification of *xylA* has been identified in the form of tandem arrays and was further confirmed experimentally by PCR and southern blot hybridization experiments. The amplification of *xylA* was linked to the strong increase in XI activity, which in turn conferred to

the strain a strongly improved D-xylose fermentation capacity. In addition, we were able to trace the step in the adaptive evolutionary process, where the amplification of *xylA* started. Interestingly an extrachromosal circular DNA plasmid (eccDNA) carrying the *xylA* gene had been generated after two transfer steps in D-xylose medium during the adaptive evolution. Further adaptation steps resulted in chromosomal integration of (part of) this plasmid in the form of a tandem repeat. As a consequence, the high D-xylose fermentation capacity was completely stable after more than 50 generations without selection pressure (chapter 2). The integration of the eccDNA into the genome might have involved homologous recombination.

5.2 Whole genome sequence analysis

The best evolved D-xylose fermenting strain GS1.11-26 and the original parent strain HDY.GUF5 have been sequenced using the Illumina next generation sequencing methodology (BGI, China). About 6 million reads from each strain were generated with an average read length of 90bp. The sequence reads obtained from BGI were mapped against the reference S288c genome sequence using CLC genomic workbench (CLC bio, Denmark). The reference sequence of chromosome XV from S288c was modified by inserting the sequence of the D-xylose/L-arabinose metabolism cassette at the *PYK2* locus, where it was originally integrated in the parent HDY.GUF5 strain. This sequence was then used as a reference together with the reference sequence of the other chromosomes. When the reads were mapped against the reference genome, more than 98% of the reads from both the parent and the evolved strain aligned to the reference sequence, with an average coverage depth of 44.

5.2.1 Detection and evaluation of copy-number variations (CNVs)

Since the coverage depth of sequence reads clearly illustrates CNVs among genomes of different strains (Araya *et al.,* 2010), the chromosomal CNVs were estimated from the depth of the whole genome sequence coverage data. At chromosomal level, the average coverage of most of the chromosomes was similar in both the evolved and the parent strains. However, chromosome IX and XVI showed 50% higher coverage in the evolved strain compared to the parent strain, while chromosome XII showed significantly reduced coverage in the evolved strain (Appendix I). To distinguish between whole chromosome duplication and segmental amplifications, and to evaluate small duplications or deletions, the sequence coverage of all 16 nuclear chromosomes was analyzed at the nucleotide level with an average window of 500bp. The log2 ratio of the read depth between the evolved and the parent strains was then calculated (Figure 5.1, A). Accordingly, we obtained a uniform coverage throughout the whole chromosome IX and XVI, at 50% higher than average in the evolved strain compared to the parent strain. The log2 ratio of 0.5 along the two chromosomes also indicated that a duplication of both chromosomes had occurred.

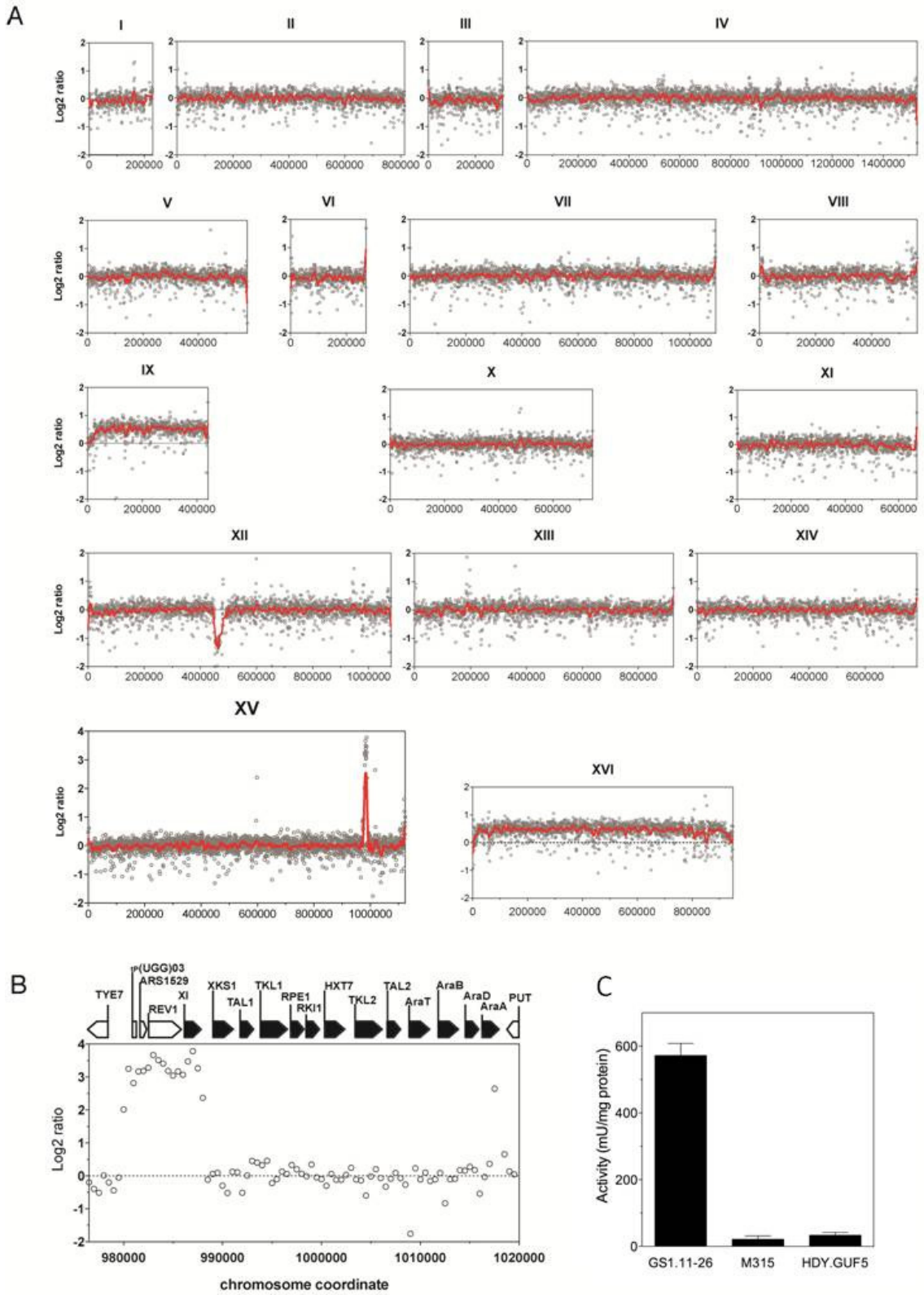


Figure 5.1. Comparison of the genome sequence coverage and XI activity between the parent strain HDY.GUF5 and the evolved strain GS1.11-26 (A) Log2 ratio depicted from whole genome sequence coverage between the evolved and the parental strain is presented for each of the 16 chromosomes. Each grey circle represents the value

of the log2 ratio obtained from sequence coverage calculated for averaged sliding windows of 500 nucleotide positions. The red line indicates the smoother trend calculated by moving average values of 10,000 bp. (B) Log2 ratio of sequence coverage between the evolved and parent strain at the *PYK2* locus of chromosome XV, where the D-xylose and arabinose gene cassette has been integrated. Annotations present in the locus are indicated by bars at the top of the figure. Bars shaded in black correspond to the heterologous genes that are inserted into the chromosome, while the unshaded bars represent part of the original yeast chromosome. The coverage was computed at individual base pair level and each circle represents the average for every 100 pb. (C) Comparison of XI activity in the parent HDY.GUF5, the mutant M315 and the evolved GS1.11-26 strain. Error bars represent the standard deviation from the mean of triplicate experiments.

In chromosome XII, the coverage at the ribosomal DNA locus has been reduced in the evolved strain by about 25% relative to the parent strain, indicating the loss of some copies of the ribosomal DNA (rDNA) genes. rDNA genes are present in about 150 to 200 tandem copies in the *S. cerevisiae* genome (Petes, 1980). The possible effect of the reduction of the rDNA copies in our evolved strain was not investigated further. However, large deletion of multiple rDNA copies is a common spontaneous phenomenon in the yeast genome. Strains with up to 50% reduction in the number of rDNA copies compared to wild type laboratory strains have no detectable defect, neither in mitotic growth nor in meiotic reproduction (Michel *et al.,* 2005).

The most prominent CNV occurred in chromosome XV where the D-xylose and L-arabinose gene cassette had been integrated. In this region, part of the integrated gene cassette, notably the *xylA* gene, and an upstream sequence that includes the genes *REV1*, a tRNA gene *tP(UUG)O3* and an autonomously replicating sequence *ARS1529,* was amplified about 9 fold (estimated from the log2 ratio) in the evolved strain compared to the parent strain (Figure 5.1, A and B). Since XI is the rate-limiting enzyme in D-xylose metabolism, this region was further investigated in more detail.

5.2.1.1 Increased Xylose Isomerase activity in GS1.11-26

In order to determine if the amplification of the *xylA* gene also correlated with higher xylose isomerase activity, we compared the activity of xylose isomerase in the evolved GS1.11-26 strain with that in the parent HDY-GUF5 and the mutant M315 strain. The GS1.11-26 strain demonstrated much higher (about 17 fold) XI activity than the parent or the mutant M315 strain (figure 5.1, C). The high XI activity is consistent with the high copy number of *xylA* in the evolved strain, though the fold increase in the XI activity was higher than that of the copy number of the *xylA* gene.

5.2.1.2 Amplification of the *xylA*-locus as tandem repeat in GS1.11-26

We sought to understand the structural arrangement of the amplification of the *xylA*-locus. Previous studies have reported amplification of genes as tandem repeat or by formation of extra-chromosomal circular DNA (eccDNA) in strains obtained by adaptive evolutionary

engineering (Dunham *et al.,* 2002); (Gresham *et al.,* 2010; Zhou *et al.,* 2012). At first notion, the presence of an autonomous replication sequence ARS1529 in the amplified *xylA*-locus (Figure 5.1, B), made us consider the possibility that this region got amplified as a self replicating eccDNA. This idea was supported by the observation of break points on either end of the amplified region when the Illumina sequence reads were mapped to the reference genome (Figure 5.2, A). The sequence reads at one end of the break point contained partially unmatched sequences that matched with the sequence of the opposite end. This condition implies either circular DNA or tandem repeat formation.

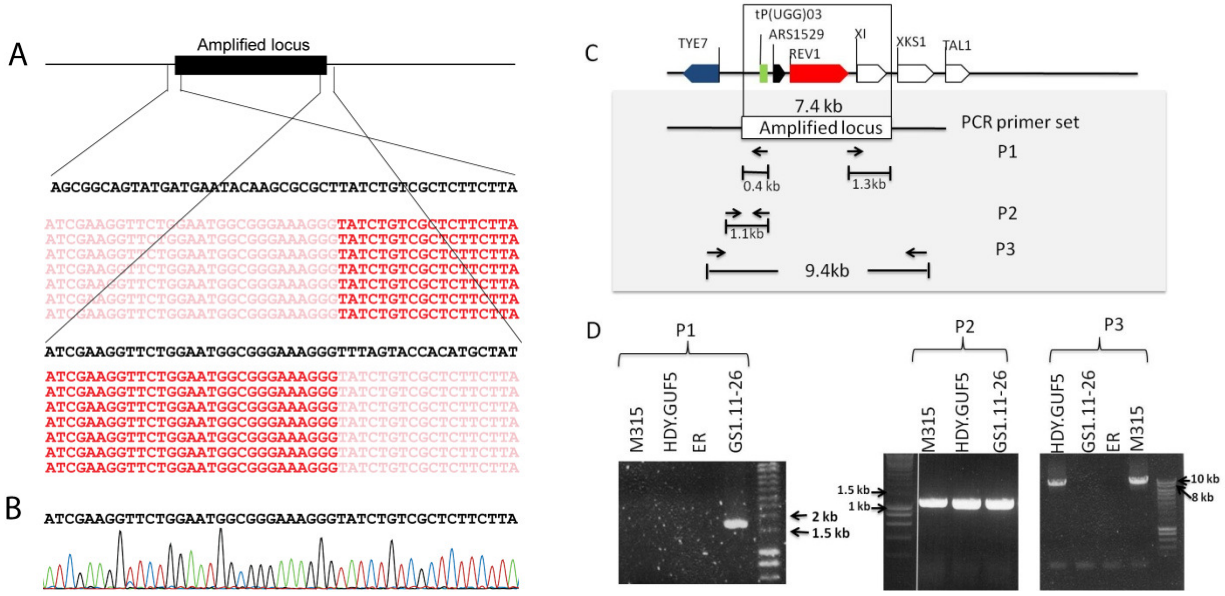


Figure 5.2. Sequence analysis at the amplified *xylA*-locus, and verification of the presence of circular or tandem repeats. (A) Illumina sequence reads mapped to the reference sequence at the amplified locus at either end. The reads that could be mapped to the reference sequence are shown in dark red, while sub-reads that do not match to the reference sequence are indicated in faded red color. The reference sequence is shown in black at the top of the reads. (B) Chromatogram from Sanger sequencing of the amplified locus, showing continuous reads through the break point of the Illumina sequence reads alignment. The Sanger sequencing was performed using a PCR product obtained by amplification with primers that anneal outwards from both ends of the locus. A similar result was obtained when the plasmid isolated from strain GS1.2-6 was sequenced. (C) Schematic representation of the amplified *xylA*-locus. Horizontal arrows stand for annealing sites and direction of the PCR primers used to verify circular or tandem repeat formation (primer set P1), the presence of the *xylA* locus at the right position (primer set P2) and single copy of the *xylA* locus (P3). The size of the expected PCR product is given in Kb. (D) Agarose gel electrophoresis picture of the PCR products obtained using the three primers sets (P1, P2 and P3) with DNA samples from different strains. ER, Ethanol Red.

To validate this, we performed PCR using a primer set P1, which consists of a pair of primers inside the amplified region that project outwards in opposite direction (Figure 5.2, C). A PCR product of 1.7kb was expected if a circular or tandem repeat sequence had been generated. The evolved strain GS1.11-26 gave a PCR product of the expected size, while no PCR product was obtained from the parent strain HDY.GUF5, the mutant M315 and the original industrial

strain Ethanol Red (figure 5.2 D). The PCR product obtained was then sequenced using conventional Sanger sequencing. The resulting contigs were shown to read-through the break point that was obtained from the alignment of Illumina sequence reads at either end of the amplified region (figure 5.2, B), indicating a continuous DNA sequence rather than a break point, which in turn points to either a circular form or a tandem repeat.

To differentiate between circular DNA and tandem repeat, we first evaluated if the chromosomal copy of the amplified region had been deleted in the evolved strain. Deletion of the locus would contradict the possibility of tandem amplification of the locus. For this purpose, two sets of primers were used (figure 5.2 C). The primer set P2 contained a forward primer annealing upstream of the amplified locus and a reverse primer annealing inside the amplified locus, and detects the presence of the *xylA* locus at the right position in the chromosome. The primer set P3, contained a forward primer upstream of the amplified locus and a reverse primer downstream of the amplified locus, and was expected to give a PCR product only upon deletion of the *xylA* locus since the locus is too big to be amplified with the PCR conditions used. A band of the expected 1.1kb size with the PCR set P2 (figure 5.2, D) and a negative result with the PCR set P3 (data not shown) was obtained in both the parent HDY.GUF5 and the evolved GS1.11-26 strain. The positive band with PCR set P2 was expected since the whole genome sequencing data indicated that at least one of the alleles was present. However, the absence of a PCR product using primer set P3 indicated that both alleles of the chromosomal locus were not deleted. Therefore, neither tandem amplification nor eccDNA could be excluded on this basis.

We then performed PCR using primer set P3 and conditions that should allow amplification of the whole amplified *xylA*-locus. A single copy of the *xylA*-locus in the genome was expected to produce a 9.4kb PCR product while chromosomal amplification of the locus should not produce any PCR product since it would be too large to be amplified. The parent HDY.GUF5 and the mutant M315 strains gave rise to a PCR product with the correct size but the evolved strain GS1.11-26 did not give rise to any band after several attempts (Figure 5.2, D). The HDY.GUF5 positive control gave rise to the expected band in all the repetitions. This indicates that only one copy of the *xylA* locus is present in each of the two alleles in the parent strain, but that the evolved strain might have multiple copies in both alleles.

To confirm this, southern blot analysis was performed with genomic DNA digested with two different restriction enzymes. First, the DNA was digested with *HindIII* that cuts only once inside the amplified *xylA*-locus. A unique probe that hybridizes in the *xylA* sequence was used to visualize the band. In the presence of only a single copy of the *xylA*-locus , a single band of 4.3 kb was expected while circular or tandem repeat sequence should give two bands of 4.3kb and 7.4 kb (Figure 5.3 A). As expected, two bands were detected for the evolved GS1.11-26 while

only one band was detected for the parent HDY.GUF5 and the mutant M315 (Figure 5.3 B). No band was detected in the negative control Ethanol Red, that does not contain the cassette in the genome.

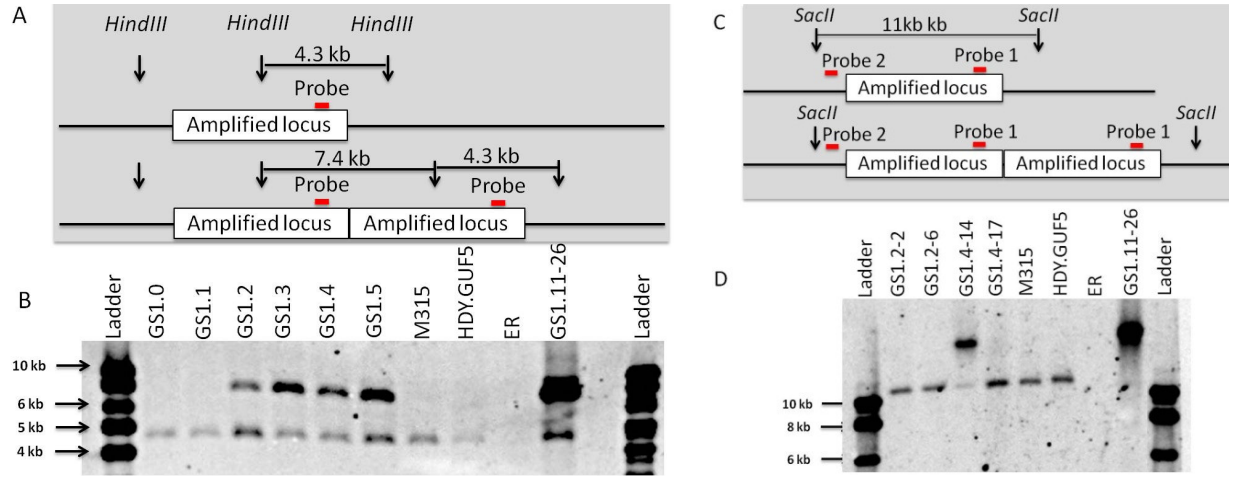


Figure 5.3. Evaluation of the amplified locus by southern blot analysis. (A) and (C) show a schematic representation of the amplified *xylA*-locus in strain GS1.11-26. Vertical arrows represent cutting sites of the restriction enzymes; the horizontal bars indicate the locus of the unique site for probe hybridization. (B) Image of the southern blot analysis after *HindIII* digestion. Two bands of expected size 4.3 kb and 7.4 kb were detected in GS1.11-26 and in cultures obtained from the second step of evolutionary adaptation onwards (GS1.2 to GS1.5). The parent strain HDY.GUF5, the mutant M315 and the culture from GS1.0 and GS1.1 showed only one band representing a single copy of the locus. The negative control Ethanol Red (ER), which does not have the gene cassette, showed no signal. (D) Southern blot image after digestion with *SacII*, which cuts only outside the amplified *xylA*-locus. Two single cell isolates from GS1.2 (GS1.2-2 and GS1.2-6) and two from GS1.4 (GS1.4-14 and GS1.4-17) were evaluated. The band of about 11kb represents a single copy of the *xylA* locus, and is present in all the tested strains except the final evolved strain GS1.11-26, which showed only a high molecular weight band. GS1.4-14 showed both the 11kb and a higher molecular weight band.

We then digested the genomic DNA with *SacII*, which cuts only outside the amplified *xylA*-locus, and hybridized with two different probes annealing either inside (same as the previous probe, Figure 5.3 C), or outside the amplified locus (between the left *SacII* restriction site and the amplified locus). An 11 kb band was expected if a single copy of the *xylA*-locus was present. The presence of the expected 11 kb fragment in the strains HDY.GUF5 and M315 using both probes hybridizing inside (figure 5.3 D) or outside the the amplified *xylA*-locus (data not shown) confirmed the presence of a single copy of the *xylA*-locus. On the other hand, the evolved strain showed only a higher molecular weight band, both with the inside (figure 5.3D) and outside probe (data not shown) confirming the presence of multiple copies of the *xylA*-locus in both chromosomal alleles. This result together with the PCR amplification using PCR set P1 (primers directed outwards on either side of the amplified locus) clearly indicates that the amplification of the locus in GS1.11-26 is in the form of a tandem repeat.

5.2.1.3 Amplification of the *xylA*-locus during evolutionary adaptation with an eccDNA intermediate

To verify at which stage of the evolutionary engineering process the amplification of the *xylA* locus had occurred, selected cultures that were isolated and kept frozen from each step of the evolutionary adaptation process were analyzed using the primer set P1, which was used to amplify the circular/tandemly repeated DNA. The cultures isolated immediately after the genome shuffling (GS1.0), and at the end of the first 5 transfers during the evolutionary adaptation (GS1.1, GS1.2, GS1.3, GS1.4 and GS1.5) were tested by PCR for the presence of the tandem amplification or circular DNA formation using PCR primer set P1. A positive PCR result was obtained for all the samples derived from the second culture (GS1.2) onwards, whereas isolates from GS1.0, GS1.1 and the original strains used for the genome shuffling step did not give rise to the PCR product (Figure 5.4). Southern blot analysis of the same samples after *HindIII* digestion also confirmed the presence of either circular or multiple copies of the locus in the samples obtained from GS1.2 onwards (Figure 5.3 B).

This strongly suggests that the amplification of the *xylA*-locus has occurred at the second step of the evolutionary adaptation process (GS1.2). As we anticipated, the sharp rise in the CO_2 production during the second culture (Figure 2.3 , chapter 2) was apparently correlated with the amplification of the *xylA*-locus. Although *xylA* was expressed from a strong promoter in the parent strain HDY.GUF5, its expression was not high enough to render strong D-xylose fermentation capacity. Amplification of the gene likely increased the expression of XI, which in turn alleviated the rate limiting bottleneck for fermentation of D-xylose.

Interestingly, the chromosomal tandem amplification of the *xylA* locus in GS1.11-26 was not detected in two D-xylose fermenting single cell clones obtained from GS1.2 (second culture). When the southern blot was performed after *SacII* digestion on these single cell isolates called GS1.2-2 and GS1.2-6, only the 11kb band was obtained for both strains, indicating that only a single copy of the gene was present in the chromosomal locus (Figure 5.3 D). This was supported by the result of the PCR amplification of the whole amplified *xylA*-locus using primer set P3, which gave the expected 9.4kb band (Figure 5.4). Since no smaller PCR band was obtained when this primer set was used (indicating that the *xylA*-locus was not deleted), and only a single band was obtained with the southern blot assay, both chromosomal alleles should have a single copy of the *xylA*-locus in these two strains. These results, together with the positive PCR result obtained using primer set P1 (Figure 5.4) in all the cultures obtained from GS1.2 onwards, clearly indicate that eccDNA was generated at the second stage (GS1.2) of the evolutionary adaptation process.

We also performed the southern blot assay with *SacII* digested DNA using genomic DNA of two single cell isolates from GS1.4 (4th culture) to test for the presence of the eccDNA. The first isolate GS1.4-14 had the highest D-xylose fermentation rate among all the isolates obtained from the culture GS1.4. Another isolate GS1.4-17 (which has only moderate D-xylose fermentation capacity) was also used for comparison. Accordingly, GS1.4-14 showed both the 11kb and a higher molecular weight band. Together with the result of the PCR assay shown in figure 5.4, this result clearly indicates the presence of multiple copies of the locus in one of the alleles and a single copy in the other allele in strain GS1.4-14. The strain GS1.4-17 showed only the 11kb band, which was also obtained by PCR amplification of the whole *xylA* locus using primer set P3 (Figure 5.4), indicating the presence of a single copy of the *xylA*-locus. This result also suggested a correlation between the amplification of the *xylA*-locus in the genome and the faster D-xylose fermentation.

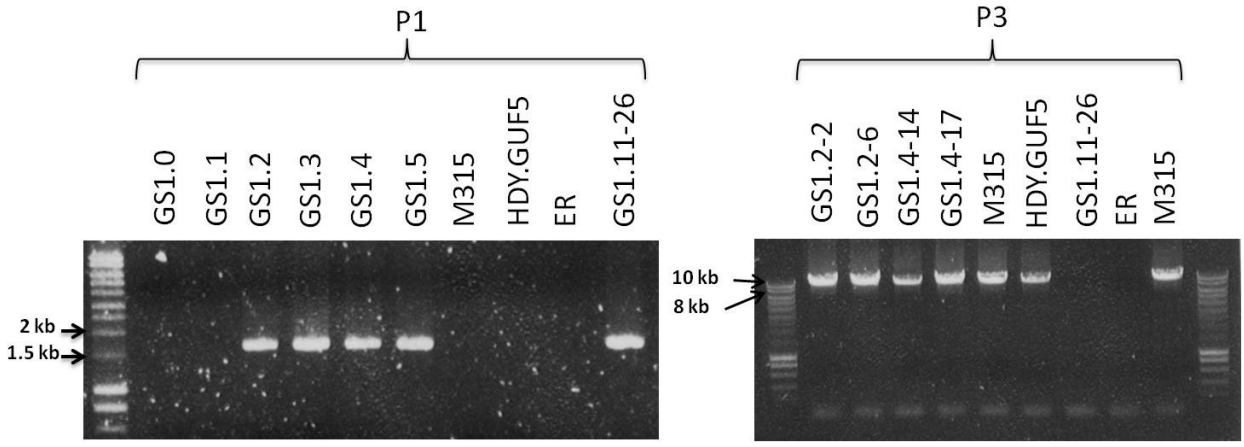


Figure 5.4. Evaluation of the *xylA*-locus in cultures and single cell clones obtained from the various stages of the evolutionary adaptation process. Agarose gel electrophoresis of PCR products obtained using primer sets P1 and P2 are shown (see Figure 5.2 C for the primer sets). DNA was isolated from a sample of the whole culture obtained just after the genome shuffling step (GS1.0) and after each batch of the first 5 cultures of the evolutionary adaptation process (GS1.1 to GS1.5). GS1.2-2 and GS1.2-6 are single cell isolates from the culture GS1.2, while GS1.4-14 and GS1.4-17 are single cell isolates from the culture GS1.4. The strains M315, HDY.GUF5, Ethanol Red (ER) and GS1.11-26 have been included for comparison.

Therefore, we have clear indications that the amplification has arisen through a circular intermediate in an early stage of the evolutionary adaptation process, and subsequently recombined in tandem array into the same locus in one of the chromosomal alleles. Later, unequal crossover or other mechanisms might have led the tandem array to be copied into the second chromosomal allele, since GS1.11-26 carried the amplified locus in both alleles.

5.2.1.4 Stability of eccDNA in intermediate strain GS1.2-6

The stability of the phenotype in GS1.2-6 was also evaluated. If the strain GS1.2-6 carried only the circular plasmid, the loss of the plasmid and therefore, loss of its ability to grow in D-xylose

medium was anticipated. To evaluate this possibility, both GS1.2-6 and GS1.11-26 were grown serially in YPD medium for about 25 generations. Two single cell isolates from each strain were then evaluated for growth in YP medium containing D-xylose as a sole carbon source. For comparison, the original strains that were pre-grown in D-xylose medium were included. Both single cell clones of GS1.2-6, obtained after about 25 generations in YPD medium, lost their growth ability in D-xylose. However, the control samples that were not exposed to YPD (pregrown in YPX) maintained the high growth capacity in D-xylose medium (Figure 5.5). Strain GS1.11-26 did not show any difference whether it was serially transferred in YPD or pregrown in YPX medium. This strongly suggested the loss of the *xylA* carrying circular DNA from GS1.2-6. Consequently, we concluded that the stability of the phenotype in GS1.11-26 is due to the integration of the circular DNA into the genome.

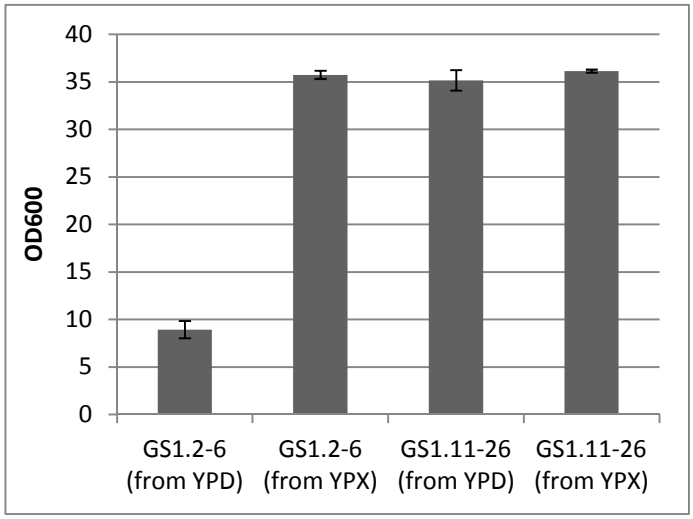


Figure 5.5. Evaluation of the stability of the *xylA*-carrying plasmid in GS1.2-6. Both GS1.2-6 and GS1.11-26 were grown in YPD medium for 5 consecutive transfers and spread for single colonies in YPD. Two single cell isolates from the serially transferred culture were tested for growth performance in YPX medium. For comparison, the original GS1.2-6 and GS1.11-26 strains were used (the pregrowth was in YPX medium for 48 h). The cells were inoculated into YPX medium at an initial OD_{600} of 0.5. Optical density was measured after 48 h at 30°C. Error bars represent standard error of the mean of the values obtained from duplicate colonies of each strain.

5.2.1.5 Isolation of eccDNA from intermediate strain GS1.2-6

To further confirm the formation of eccDNA, plasmid DNA isolation was performed from the strain GS1.2-6 and GS1.11-26, using a modified protocol from (Singh and Anthony Weil, 2002). Cells were pre-grown in 100 ml YPX medium for 24 h to enrich the plasmid content in the cells. The whole 100 ml culture was used for plasmid isolation. Using this procedure, a substantial amount of plasmid DNA (more than 200 ng/µl) was obtained from GS1.2-6 (figure 5.6). However, the amount of DNA obtained from GS1.11-26 was too low to be conclusive (data not shown). This is probably due to the loss of the plasmid in the GS1.11-26 strain during the

subsequent evolutionary steps, since there is no need for the strain to maintain the plasmid when enough copies of the essential gene *xylA* have been integrated in the genome to sustain fast D-xylose utilization.

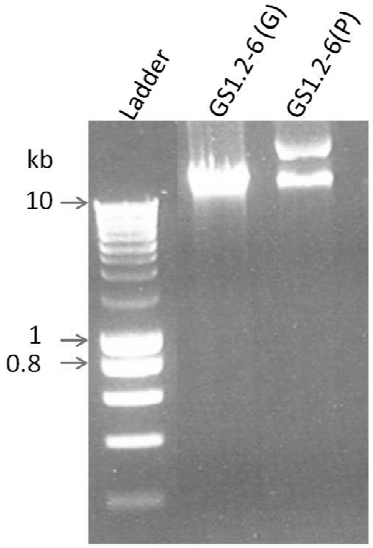


Figure 5.6. Plasmid DNA isolated from strain GS1.2-6. Plasmid DNA was isolated from a sample of the strain GS1.2-6. Separation was done using 0.8% agarose gel electrophoresis. (P) stands for plasmid DNA while (G) stands for genomic DNA, which was used for comparison. Two bands in the GS1.2-6 (P) lane might represent different structural forms of the plasmid, probably supercoiled (lower band) and nicked (upper band).

5.2.1.6 ARS1529 is a functional origin of replication in *S. cerevisiae* strain

ARS1529 was previously shown to be a functional replication site in yeast (Nieduszynski *et al.,* 2006). In order to validate the functionality of the ARS1529 in the eccDNA intermediate, we assessed the self replication ability of this sequence since there were 5bp deletions and 5 SNPs in an AT-rich region just 13bp downstream of the ARS consensus sequence (ACS), compared to the reference genomic sequence of S288c (Figure 5.7). We PCR amplified the region containing ARS1529 together with the tRNA coding sequence and the *xylA* gene (*xylA*-ARS) from the genomic DNA of the evolved strain GS1.11-26 (figure 5.2C) and cloned the PCR product into a yeast integrative vector containing kanMx gene as a selection marker. After transformation of this construct into the mutant M315, we obtained several transformants with similar transformation efficiency as that of the 2µ-based plasmid, showing that the plasmid was able to propagate via ARS1529. Loss of the plasmid was observed after some generations in non-selective conditions indicating that the plasmid was not integrated into the genome. This confirmed that the ARS1529 sequence can function as replication origin in *S. cerevisiae.*

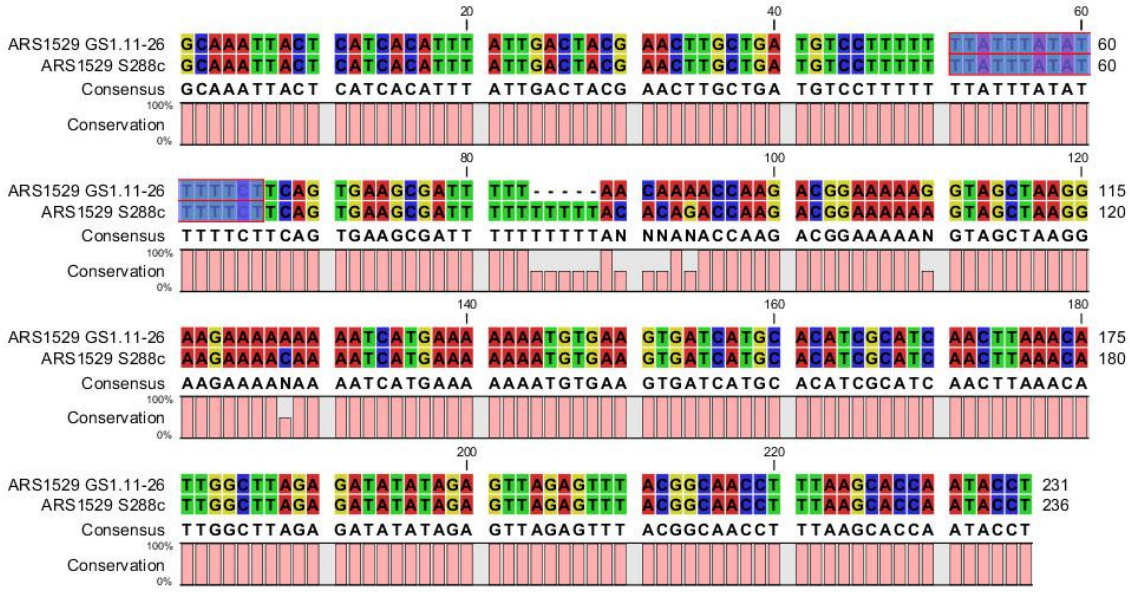


Figure 5.7. ARS1529 sequence comparison between the evolved strain, GS1.11-26, and the reference genomic sequence of S288c. Shaded blue region represent the ARS consensus sequence (Nieduszynski *et al.,* 2006). A gap of 5 base pairs in GS1.11-26 from 84th to 88th bp is shown with dashed lines. The sequence from GS1.11-26 was confirmed by Sanger sequencing and was identical to the sequence obtained by Illumina sequencing.

5.2.1.7 Amplification of *xylA* is not the sole reason for the high D-xylose fermentation capacity

In order to determine whether amplification of the *xylA* gene is the only reason for the high D-xylose fermentation performance of GS1.11-26, and also whether it is still a limiting factor for D-xylose fermentation capacity in that strain, the *xylA*-ARS plasmid described above and a 2-micron based plasmid containing the same *xylA* gene, were transformed into the mutant M315, the evolved strain, GS1.11-26, and the parent strain, HDY-GUF5. When the transformants were tested for D-xylose fermentation, there was no further improvement of GS1.11-26 upon introduction of the plasmids, indicating that XI might not be a limiting factor anymore in GS1.11-26 under the fermentation conditions used (Figure 5.8). In addition, evaluation of several transformants of the HDY-GUF5 strain with both ARS based and 2µ based plasmids carrying the *xylA* gene did not bring about any notable improvement of D-xylose fermentation capacity, showing that the amplification of *xylA* is not the sole reason for the superior D-xylose fermentation phenotype of GS1.11-26. On the other hand, when either plasmid was transformed into the mutant strain M315, the transformants showed much higher D-xylose fermentation capacity, though not as high as GS1.11-26 (Figure 5.8). This shows that M315 has one or more crucial mutations in the genome (generated by the mutagenesis step) that are essential for the superior D-xylose fermentation phenotype in combination with over-expression of XI.

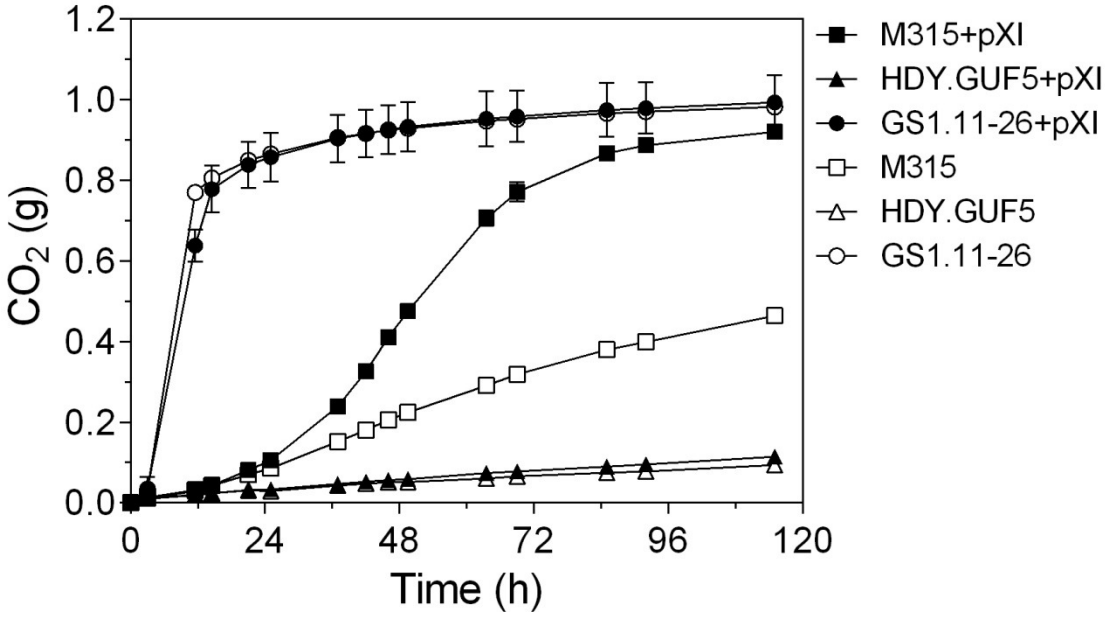


Figure 5.8. Evaluation of XI over-expression on D-xylose fermentation capacity through expression of the *xylA* gene from a multi-copy plasmid. pXI stands for multicopy plasmid containing *xylA*. Only the 2µ based plasmid is shown. The ARS based plasmid gave a similar result. Each strain was inoculated into YP medium containing 40 g/L D-xylose, at an initial cell density of 1.3 gDW/L and incubated at 35 °C. The CO_2 production was estimated from the weight loss of the total culture. Error bars represent standard error of the mean between duplicate experiments (using two independent transformants).

In addition, other beneficial genetic changes might have accumulated during the subsequent genome shuffling and/or evolutionary engineering process. This can explain why the M315 strain, containing either the ARS based or the 2µ based XI plasmid, did not ferment D-xylose as good as the final strain GS1.11-26. However, the difference in the phenotype might also be due to the difference in the expression level of XI, since the copy number of the plasmids is unknown and may be lower than the total amplification of *xylA* in the genome.

5.3 Detection of Single nucleotide polymorphism (SNP)

5.3.1.1 Comparison of SNPs between the parent HDY.GUF5 and the evolved GS1.11-26 strains

The whole genome sequence of the evolved and the parent strain has also been compared to find unique SNPs between the strains that might play a role in establishing the superior phenotype. After the genome sequence reads from each strain were mapped against the reference sequence, SNPs relative to the reference strain were first computed using Lasergene's SeqMan Pro software (DNASTAR). Compared to the reference laboratory strain, there were 55,465 and 49,458 SNPs present in the parent and the evolved strain, respectively. Most of the SNPs were homozygous (90% in both strains) with an overall heterozygosity rate of 0.056 % and 0.051 % for the parent and evolved strains, respectively. Since the majority of published genome sequences are from haploid yeast strains, a proper comparison of the heterozygosity level was not possible. However, the level of heterozygosity obtained in our

study was in the range of the values reported in a recently published work on whole genome sequence analysis of industrial strains (Borneman *et al.,* 2011).

To compare the nucleotide variants among the parent HDY.GUF5 strain and the evolved GS1.11-26 strain, the SNPs obtained were sorted for uniqueness in either of the two strains in comparison with the reference genome sequence of S288c. This resulted in 2304 and 1992 SNPs, which are unique to the evolved strain and the parent strain, respectively, compared to the S288c sequence. The 1992 SNPs unique to the parent strain might indicate the reversion of SNPs, during the mutagenesis, genome shuffling and evolutionary engineering of the parent strain towards the evolved strain, to the allele of the reference S288c sequence. This means that the final strain, GS1.11-26, must have undergone 4296 mutations during the mutagenesis, genome shuffling and evolutionary engineering of the parent strain, 1992 mutations back to the reference sequence and 2304 new mutations.

Table 3.1.Comparison of nucleotide variant statistics for the sequence of the evolved strain GS1.11-26 and the parent strain HDYGUF5, compared to the reference sequence of S288c. Heterozygosity rate stands for the percentage of genome with heterozygous nucleotide variants.

Nucleotide variants *	HDY.GUF5	GS1.11-26
Total SNPs	55465	49458
Total small Indels	2139	1762
Heterozygous SNPs	5604	5295
Heterozygous small Indels	1192	914
Heterozygosity rate (%)	0.056	0.051

* SNPs and small Indels that are outside high or low coverage regions were considered.

5.3.1.2 Evaluation of a *TKL2* mutation in GS1.11-26

In order to evaluate SNPs that might be linked to the phenotype, unique non-synonymous mutations in genes, and mutations that were 500 bp upstream or 300bp downstream of an ORF were identified. Since there were more than 700 such SNPs present in the evolved strain relative to the parent strain, we selected genetic variations in genes that might be related to sugar metabolism and genes that were previously reported to be linked with D-xylose fermentation capacity, for further experimental evaluation.

The best candidate SNP was the homozygous non-synonymous mutation (G102R) that occurred in the coding region of the codon optimized *TKL2* (opt_*TKL2*) gene located inside the integrated D-xylose/L-arabinose gene cassette. Since transketolase is one of the key enzymes of the non-oxidative PPP, and was shown to be important for faster D-xylose fermentation (Karhumaa *et al.,* 2005), we analyzed if the mutation in the opt_*TKL2* was important for the high rate of D-xylose fermentation in the evolved strain.

5.3.1.2.1 Over-expression of the mutant *TKL2* allele (opt_*tkl2*^{G102R})

We first evaluated whether over-expression of the mutant allele improves D-xylose fermentation in the parent strain. For that purpose, the opt_*tkl2*^{G102R} allele was amplified from GS1.11-26 and cloned into the plasmid pXI to generate pXI_opt_*tkl2*^{G102R}. The plasmid pXI containing the *xylA* gene was used for the cloning since over-expression of *xylA* is an essential requirement for D-xylose fermentation. This new plasmid (pXI_opt_*tkl2*^{G102R}) was then transformed into the strains HDY-GUF5, GS1.11-26 and M315.

However, the transformants of strain HDY.GUF5 expressing pXI_opt_*tkl2*^{G102R} did not show any improvement in D-xylose fermentation capacity (Figure 5.9 A). In addition, M315 expressing pXI_opt_*tkl2*^{G102R} showed a reduced rate of D-xylose fermentation compared to strains expressing only pXI. This is apparently due to the undesired over-expression of the mutant *TLK2* allele from the multi-copy plasmid. A similar reduced rate of D-xylose fermentation associated with overexpression of *TKL2* has been shown in a previous study (Matsushika *et al.,* 2012). Although the *TKL2* gene over-expressed in our work was modified, it also resulted in inhibition. On the other hand, such a reduced rate of D-xylose fermentation was not observed in GS1.11-26 expressing the plasmid pXI_opt_*tkl2*^{G102R}. Since plasmid copy number inversely correlates to its size (Zhong *et al.,* 2011) we hypothesized that the larger size of pXI_opt_*tkl2*^{G102R} compared to pXI might result in a reduced copy number. Since a high copy number of *xylA* is essential for a high D-xylose fermentation rate, the lower copy number of pXI_opt_*tkl2*^{G102R} might be the reason for reduced rate of D-xylose fermentation in pXI_opt_*tkl2*^{G102R} expressing strains of M315 compared to those expressing only pXI. This problem was probably not manifested by GS1.11-26, because of the presence of multiple copies of the *xylA* gene in the genome (Figue 5.1).

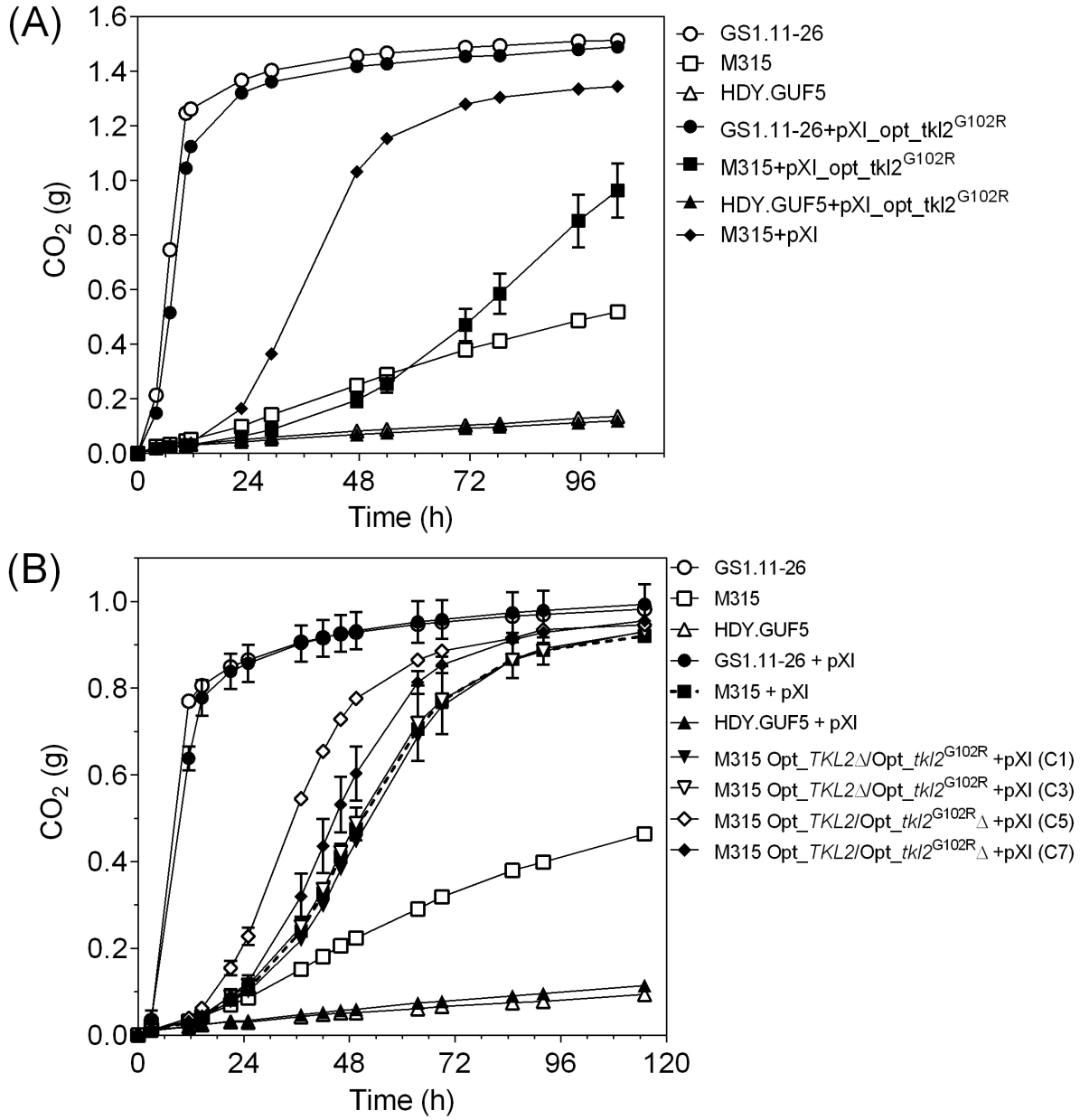


Figure 5.9. Evaluation of the effect of a SNP in the opt*TKL2* allele in strain M315 expressing pXI. (A) Comparison of D-xylose fermentation performance of three strains M315, HDY.GUF5 and GS1.11-26 carrying the plasmid pXI_opt_*tkl2*^{G102R} with strains that do not contain the plasmid. Error bars represent standard error of the mean for the result from at least two independent transformants. (B) Two independent single opt_*TKL2* deletion mutants of M315 that carried only the opt_*tkl2*^{G102R} allele (strains C1 and C3), and two other single deletion mutants carrying the wild type opt_*TKL2* allele (strains C5 and C7) were transformed with the pXI plasmid. The transformants were evaluated for fermentation performance in YP+4% D-xylose at 35°C with an initial cell density of OD 5. Error bars represent standard error of the mean of the results obtained with two independent colonies transformed with pXI. Strain M315+pXI is shown in broken line for clear comparison.

5.3.1.2.2 The mutant opt_*TKL2* allele (opt_*tkl2*^{G102R}) negatively influences D-xylose fermentation relative to that of the wild type opt_*TKL2* allele

We then evaluated if just the chromosomal replacement of the wild type opt_*TKL2* allele with the mutant opt_*tkl2*^{G102R} has an effect in the mutant strain M315. The strain M315 is heterozygous for the opt_*TKL2* allele, since it has already one copy of opt_*tkl2*^{G102R} and one copy of the wild type opt_*TKL2* allele. Deletion of either allele can produce hemizygous strains that differ only with these alleles, and therefore allow a good comparison of the effect of the mutation. In addition, unlike the parent strain, the mutant M315 was able to ferment D-xylose well when the multicopy plasmid containing *xylA* was introduced. Therefore, the effect of the *opt_TKL2* alleles can be easily evaluated.

For this purpose, we first amplified the bleomycin resistance gene from the plasmid *pUG66* using primers that contain a 50bp homologous sequence upstream and downstream of the promoter and terminator region, respectively, of the opt*TKL2* gene. Transformation of this PCR product into M315 resulted in single deletion of either the wild type or the mutant allele. Seven transformants were randomly selected and the region containing the mutation was sequenced to distinguish which allele was deleted. Two transformants retained the wild type allele (the mutant allele was deleted), while 4 transformants retained the mutant allele (the wild type allele was deleted). One transformant carried both alleles indicating that the marker might have integrated into a different locus, and that strain was not used.

We then expressed the plasmid pXI into two transformants that carried only the wild type opt_*TKL2* allele (opt_*TKL2/*opt_*tkl2*^{G102R}Δ +pXI) and two transformants that carried only the mutant allele (opt_TKL2Δ/opt_*tkl2*^{G102R} +pXI). The strains were then tested for fermentation performance in YP medium containing only D-xylose as a carbon source. No antibiotic was used for selection, since growth in D-xylose can be used to maintain the pXI plasmid.

As can be seen from figure 5.9 B, strains carrying only the wild type opt_*TKL2* allele showed a slightly faster rate of fermentation compared to the original M315 strain that carried both alleles. On the other hand, strains containing the mutant opt_*tkl2*^{G102R} allele alone did not show any notable difference in fermentation rate compared to the M315 strain. This appears to indicate that the mutation in the opt_*TKL2* gene might have a slight negative influence on the rate of D-xylose fermentation. Hence, it is apparently not one of the causative mutations for the high D-xylose fermentation rate of strain GS1.11-26.

5.4 Discussion

Recombinant strains of *S. cerevisiae* expressing the XI pathway displayed higher yield of ethanol per consumed D-xylose compared to strains expressing XR/XDH. However, the rate of D-xylose utilization by XI-expressing strains was found to be inferior to the rate obtained by XR/XDH expressing strains (Karhumaa *et al.,* 2007b), mostly due the low activity of the bacterial or fungal XI in *S. cerevisiae*. As a result, expression of XI was accomplished only with multicopy plasmids and strong promoters in the previously constructed strains (Kuyper *et al.,* 2005a, 2005b; Traff *et al.,* 2001). Chromosomal integration of active *Piromyces sp. xylA* gene in an *S. cerevisiae* strain did not render any D-xylose growth ability (Karhumaa *et al.,* 2007b).

In chapter 2, we reported the construction of an industrial strain through chromosomal integration of the *C. phytofermentans xylA* expressed from a strong promoter. Similar to the previous study (Karhumaa *et al.,* 2007b), the recombinant strain HDY.GUF5 in our study was unable to efficiently utilize D-xylose aerobically or anaerobically. However, the evolved strain GS1.11-26, that was derived from HDY.GUF5 by mutagenesis, genome shuffling and evolutionary engineering, displayed very efficient D-xylose utilization rate.

In this chapter, we have described the unraveling of part of the genetic basis behind the efficient D-xylose fermentation rate of GS1.11-26, by comparing the whole genome sequence of GS1.11-26 with the sequence of its parent strain HDY.GUF5. The most striking genetic difference was found in the chromosomal region where the *xylA* gene was integrated. In this region, the copy number of the *xylA* gene was amplified about 9 fold in the evolved strain GS1.11-26 compared to the parent strain. Interestingly, the amplification of *xylA* was correlated with elevated activity of XI. As described above, the inherently low activity of XI in the recombinant strains developed previously was a limiting factor for efficient D-xylose assimilation. Therefore, the high D-xylose assimilation rate by the evolved strain GS1.11-26 is due, at least in part, to the high copy number of *xylA*, which resulted in high XI activity.

A similar, elevated D-xylose utilization rate due to multicopy integration of *xylA*, has been reported recently in a strain developed by expression of the *Piromyces sp*. XI and further evolutionary adaptation (Zhou *et al.,* 2012). In that report, the original recombinant strain was constructed through expression of XI and XK from a multicopy plasmid. Since both genes coding for XI and XK were under the same (but separate) promoter and terminator sequences (TDH3p and CYC1t) it was suggested that duplication and further copy number increase of the *xylA* gene might have been initiated though unequal crossover of homologous regions in the plasmid. However, the recombined plasmid could not be isolated from the evolved strains due to the integration of the plasmid into the genome. In our study, however, the original recombinant strain HDY.GUF5 had been constructed through direct integration of the gene cassette into the

genome using an integrative vector. Moreover, we have confirmed the presence of a single copy of the *xylA* gene in the original recombinant strain HDY.GUF5. Therefore the multiplication of the *xylA* locus did not start from the original plasmid that was used to construct the recombinant strain. On the other hand, we have clearly demonstrated the occurrence of self-replicating eccDNA carrying the *xylA* gene in the course of the evolutionary process. Both PCR and southern blot assays could not detect the presence of eccDNA in the parent strain HDY.GUF5, the original mutant strain M315, the culture after the genome shuffling step (GS1.0) and the first batch culture of the evolutionary adaptation process (GS1.1). On the other hand, the presence of eccDNA was confirmed in the second culture (GS1.2) and therefore seems to have started from that point on. The dramatic increase in the rate of D-xylose fermentation during the second step of the evolutionary adaptation process was striking and can now be explained by the gain of this crucial genetic change. The formation of self-replicating eccDNA carrying the *xylA* gene alleviated the rate limiting step in the D-xylose fermentation pathway. Further evolutionary adaptation resulted in chromosomal integration in the form of a tandem repeat. As a consequence, the high D-xylose fermentation phenotype was completely stable for more than 50 generations without selection pressure.

The multiple integration of the *xylA*-locus as tandem array in the genome seems to improve the D-xylose fermentation rate compared to the eccDNA form. This was evident from the fact that GS1.2-6 containing only the eccDNA still ferments much slower than GS1.4-14, which has the tandem amplification in one of the alleles of the chromosomal locus. In addition, the strain GS1.11-26 that contains the amplified *xylA*-locus in both alleles, showed faster D-xylose fermentation performance than GS1.4-14. This shows that higher copy number of *xylA* improves the rate of D-xylose fermentation capacity. However, the possibility that other genetic changes might have arisen during the subsequent evolution process cannot be excluded.

The phenomenon of circular DNA intermediate formation and tandem integration into the genome has been suggested to have occurred in wine yeast in the course of natural evolution. The possibility of extra-chromosomal circular DNA (eccDNA) formation in that study was supported by the presence of autonomously replicating sequences at the integrated locus (Galeote *et al.,* 2011). In addition, self replicating eccDNA formation was recently reported in *S. cerevisiae* during adaptation to nitrogen starvation (Gresham *et al.,* 2010). Though formation of eccDNA frequently involves terminal repeat sequences flanking a mobile DNA region (Galeote *et al.,* 2011), there were no such repeat sequences in the parent strain HDY.GUF5 nor in the evolved strain GS1.11-26 at the *xylA*-locus, that might have resulted in homologous recombination for excision. In addition, we showed that the locus was not excised in all the

cultures tested. The mechanism of eccDNA formation is not yet clear and needs further investigation.

The effect of high copy number amplification of *xylA* was further verified through over-expression of the same *xylA* gene in the mutant strain M315. Since over-expression of the *xylA* gene in the parent strain did not confer the D-xylose fermentation phenotype, but significantly improved that of the mutant M315 strain, one or more crucial mutations have apparently been generated during the mutagenesis step of our strain improvement method. This result supports the importance of chemical mutagenesis for strain development prior to evolutionary engineering.

Several SNPs have been identified between the evolved and the parent strain. The SNP in the opt_*TKL2* gene was evaluated in the mutant strain M315 that contains heterozygous opt_*TKL2* alleles. Reciprocal deletion of either the mutant or the wild type allele revealed a slightly beneficial effect of the wild type allele compared to the mutant allele in the M315 strain over-expressing *xylA*. Therefore, replacement of the mutant alleles with the wild type allele in GS1.11-26 might be an important way to further improve the D-xylose fermentation performance of GS1.11-26. On the other hand, this result indicated that the *TKL2* mutation was unlikely to be a causative mutation for the high D-xylose fermentation rate of strain GS1.11-26.

In conclusion, we demonstrated that over-expression of *xylA* is crucial for high D-xylose fermentation capacity. However, the rapid D-xylose consumption capacity requires a synergistic interaction between high XI activity and one or more mutations in the genome. Since there were many other genetic differences between the parent and the evolved strain, evaluation of each individual genetic difference for causing the superior D-xylose fermentation phenotype would be impractical. Therefore, other more powerful methods, like pooled segregant whole-genome sequence analysis for mapping the quantitative trait loci, might be more useful to rapidly identify the mutations that are responsible for the rapid D-xylose fermentation phenotype. Additionally, the generation of eccDNA as a means of adaptation during evolution in D-xylose medium was confirmed.

Chapter 6

Genetic mapping of mutation responsible for high D-xylose fermentation rate in GS1.11-26 using pooled segregant whole genome sequence analysis

6.1 Introduction

As described in the previous chapter, identification of the genetic determinant(s) of a phenotypic trait is a crucial but also the most difficult step in inverse metabolic engineering. Using whole genome sequence comparison of the evolved strain GS1.11-26 to its parent strain HDY.GUF5, we obtained more than 4000 single nucleotide variants (SNPs) distributed over the genome (chapter 5). However, linking one or several of the thousands of genetic changes with the phenotype of improved xylose fermentation in GS1.11-26 was very difficult. Even though the amplification of the *xylA* gene in chromosome XV has been identified as one of the essential genetic changes responsible for efficient D-xylose fermentation, we also showed that other genetic factors, yet to be identified, are required for the high D-xylose fermentation capacity in GS1.11-26. This is already an indication for the polygenic nature of the trait.

Recent advances in high-throughput genotyping technologies have facilitated the rapid identification of genetic factors responsible for a phenotype of interest. Various advanced methods and genetic tools that aid the rapid identification of genetic targets for metabolic engineering have been previously described (Swinnen *et al.,* 2012b).

Among the various high-throughput genotyping technologies, pooled-segregant whole-genome sequence analysis (PSS) was shown to be a powerful and straightforward tool to map the loci with a genetic factor responsible for a quantitative trait (Swinnen *et al.,* 2012a). In this method, a superior haploid strain with a certain phenotype of interest is crossed with a genetically distinct inferior haploid strain that does not display the phenotype. Haploid segregants from the hybrid strain are screened for the trait of interest and those segregants displaying the phenotype, as good as the superior parent, are selected. Afterwards, equal amounts of cells from the selected segregants are pooled and the genomic DNA from the pool is sequenced. When the frequency of the single nucleotide polymorphism (SNP) in the pool of segregants is scored, genetic variants that are not linked to the phenotype segregate randomly with a ratio of about 0.5 and therefore are evened out. On the other hand, the regions where the genetic determinants of the phenotype are located mostly originate from the superior parent; as a result, the SNPs located closer to the locus are overrepresented in the pool showing a ratio close to 1 depending on the importance of the genetic determinant (Figure 6.1). In a control pool containing randomly selected segregants, the SNP frequency remains close to 0.5.

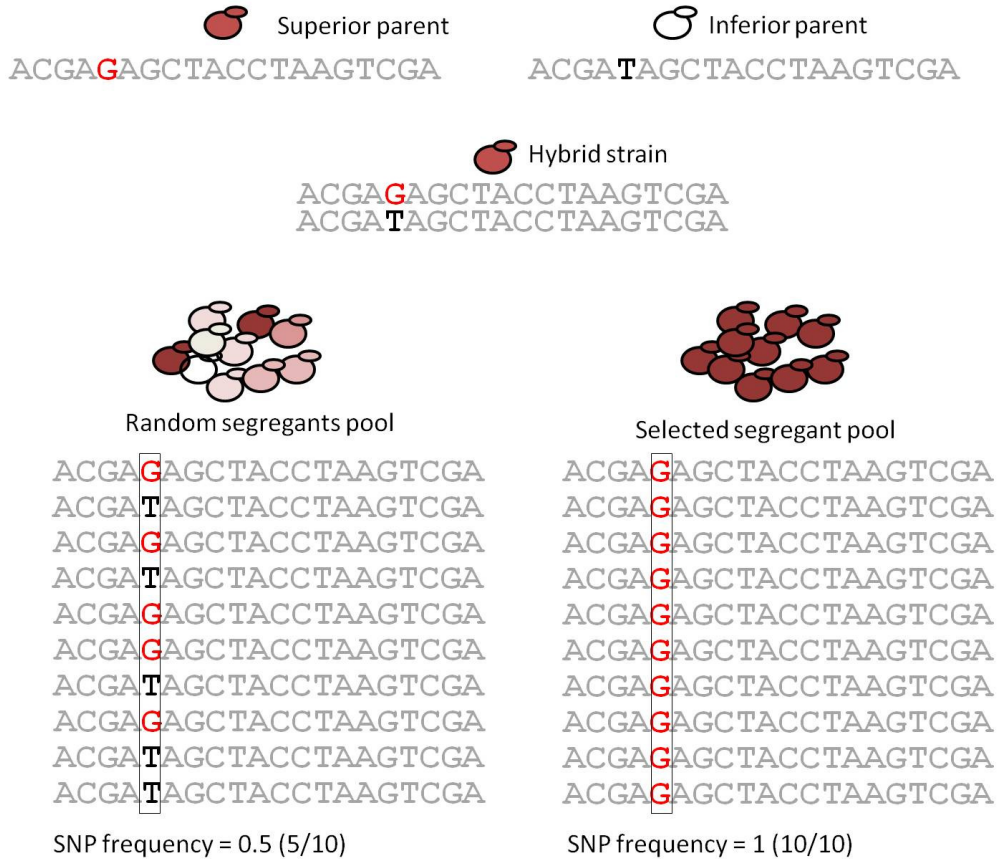


Figure 6.1. Schematic representation of QTL mapping in *S. cerevisiae* using SNPs as genetic markers. A superior strain with a trait of interest is crossed with an inferior strain without the trait. The letters in grey, red or black represent the nucleotides at a chromosomal region taken as an example. The SNP nucleotide from the superior strain is shown in red while the nucleotide at the same position in the inferior parent is shown in black. The hybrid strain obtained by crossing the two original strains is sporulated and segregants that show the trait of interest are selected. The possible SNP frequency in a pool of randomly selected segregants or a pool of selected segregants with the phenotype of interest is shown at the bottom. The SNP nucleotide from the superior parent strain is overrepresented in the pool with superior segregants, indicating that it is located in a region with a QTL, i.e. a locus with one of the genetic changes that determine the trait of interest.

One of the limitations of PSS is the requirement for haploid strains of opposite mating type as parent strains. Most industrial yeast strains are diploid, polyploid or aneuploid. Obtaining a stable haploid derivative from such industrial strains displaying the trait of interest that is as good as the parent strain, can be a strong challenge(Swinnen *et al.,* 2012b). The genomic complexity of industrial yeast strains is one of the most important limitations for application of polygenic analysis by PSS. In this chapter, we have used PSS for identification of genetic loci responsible for the high D-xylose fermentation rate in the diploid strain GS1.11-26. Unlike the regular PSS, we have mated the diploid (aneuploid) GS1.11-26 strain with a mating competent diploid (aneuploid) strain with the opposite mating type. In this way, we have evaluated the possible application of this technology directly with the parent strains, which would alleviate

the need to isolate a haploid derivative with the same trait of interest. If successful, this would have significant implications for the genetic mapping of phenotypic traits directly with industrial yeast strains.

6.2 Genetic mapping by "pooled-segregant whole genome sequence analysis" (PSS)

To perform the PSS, the D-xylose fermenting diploid *MATα/α* strain GS1.11-26 was used as the superior starting strain, and was crossed with Fseg25, a diploid *MAT**a/α*** segregant of a genetically unrelated baker's yeast that does not ferment xylose (chapter 4). We then screened 819 segregants from the tetraploid hybrid strain, and subsequently selected 27 diploid segregants that ferment D-xylose at a rate close to GS1.11-26 (the screening and isolation of the 27 best segregants is explained in chapter 4).

6.2.1 Whole genome sequencing

The pooled genomic DNA of the 27 best D-xylose fermenting segregants, a control pool of 27 segregants that do not ferment D-xylose but showed growth on solid medium with xylose, and the genomic DNA of the inferior parent Fseg25, were sequenced using high-throughput Illumina sequencing technology (BGI, China). Paired end sequencing was conducted with a 500 bp library. To ensure selection of high quality sequence reads, the raw data were filtered for several parameters including removal of adaptor contamination and low quality reads (table 6.1). We received the resulting high quality reads for further bioinformatics analysis.

Table 6.1. Statistics of the Illumina sequence reads before and after quality filtering. "Clean data" represents all the reads obtained after filtering and removal of low quality reads from the raw data (all the reads before quality filtering). Mbp, million base pair.

Sample name	Insert size(bp)	Raw data(Mbp)	Clean data (Mbp)
Superior pool	500	533	519
Inferior pool	500	532	518
Inferior parent (Fseg25)	500	520	506

6.2.2 Bioinformatics analysis

Reads obtained from each pool and from the inferior parent strain (and from GS1.11-26, which was sequenced previously, chapter 5) were mapped against the sequence of the reference laboratory strain S288c using SeqMan Ngen (Lasergene). The SNPs present in each pool and in the two parent strains (GS1.11-26 and Fseg25) were computed. Next, the SNPs in GS1.11-26 that were not shared by the inferior parent Fseg25 were determined, resulting in a set of specific SNPs between GS1.11-26 and Fseg25. Only these SNPs unique to the superior parent GS1.11-26 compared to the inferior parent Fseg25 were then selected from the SNP list of the superior and inferior pool. This is because SNPs (compared to S288c as reference) that are present in both parents are always present in all the segregants and are thus useless for genetic

mapping. Subsequently, the variant frequency of each SNP was plotted against its chromosomal position using the statistical software R. Smoothening of the SNP frequency was performed using an algorithm Linear Mixed Model (LMM), under smoothing splines (Claesen *et al.,* 2013). In this model, the scattered SNP variant frequency points are transformed into an average line over the length of the chromosome taking into account the sequence depth, the sequence quality and the distance between polymorphisms.

6.2.3 Segregation in diploid segregants (tetraploid parent)

In order to make statistical inferences for the identification of putative QTLs, we first examined the possible biological pattern of inheritance (meiotic segregation) of polymorphisms in diploid segregants obtained from a tetraploid parent. Polymorphisms in haploid segregants obtained from a diploid parent normally follow a 2:2 segregation. Therefore, in regular PSS, the average SNP variant frequency for random segregation (locus not linked to the phenotype) is around 0.5. In this case, a statistically significant deviation of the SNP variant frequency from 0.5 indicates the presence of a putative QTL (Swinnen *et al.,* 2012a).

On the other hand, segregation of spores from a tetraploid strain follows a more complex chromosomal inheritance than the regular 2:2 segregation manifested by haploid segregants from a diploid parent. The four meiotic spores of a tetrad made by a tetraploid parent mostly contain a diploid genome (Albertin *et al.,* 2009). Although the possibility of aneuploidy was not ruled out, we also showed that all the 27 superior segregants obtained from the tetraploid parent GS1.11-26/Fseg25 had a diploid genome when the DNA content was estimated by flow cytometry (chapter 4).

The polymorphisms in a tetraploid parent do not follow the regular 2:2 segregation in the diploid offspring. The chromosomes in an autotetraploid yeast (generated from two strains of the same species) have no preferential pairing during meiosis. This results in random bivalent pairing, formation of quadrivalents or a combination of both during meiosis (Stift *et al.,* 2010). Since the tetraploid parent in our study was generated by crossing two strains of the same species (autotetraploid), we assumed that the meiotic spores followed tetrasomic inheritance (mendelian or random segregation). Because the effect of the quadrivalent pairing during meiosis has relatively little influence on the SNP frequency in the segregants, tetrasomic inheritance with bivalent pairing was assumed to determine the expected SNP frequency in the segregants. An example of the SNP inheritance pattern in the segregants is given in figure 6.2.

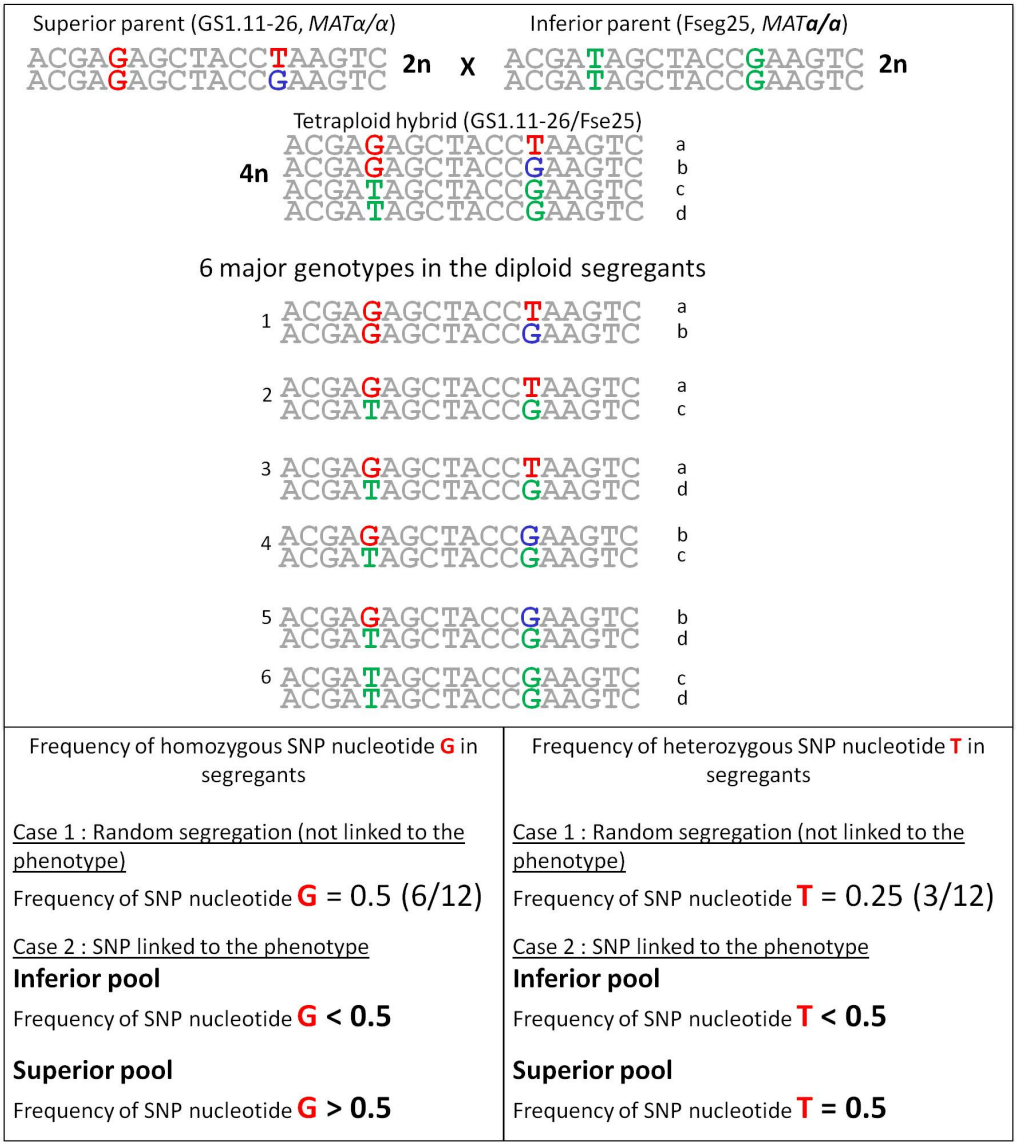


Figure 6.2. SNP inheritance and possible SNP variant frequency in diploid segregants obtained from tetraploid parent. Each sequence in the upper box represents a region of a chromosome with nucleotide polymorphisms indicated in red. Nucleotides in grey, blue or green are the same as in the reference strain. Green is used for nucleotides that originate from Fseg25, while blue is used for nucleotides from GS1.11-26. The 6 possible genotypes are listed and numbered from 1 to 6. The Mendelian random segregation with only bivalent pairing was used for calculation.The lower left box shows the possible segregation of a region with a homozygous SNP nucleotide (G in red). In case 1 with random segregation, the SNP variant frequency in the segregants will be 0.5 (6 G and 6 T). In case 2, when the homozygous SNP nucleotide (G in red) is linked to the phenotype, the SNP frequency in the inferior pool will be less than 0.5 (close to 0) since the segregants will mostly contain the wild type allele T (genotype number 6). The SNP variant frequency in the superior pool will be higher than 0.6, since all the segregants should contain this SNP nucleotide (G in red, genotypes 1 to 5). In this case, the SNP variant frequency varies depending on the type of mutation (dominant or recessive). The lower right box shows the possible frequency of a heterozygous SNP nucleotide in segregants. During random segregation (case 1), the SNP variant frequency will be 0.25 (3/12). When the SNP nucleotide is important for the phenotype (case 2), the SNP frequency will be much lower than 0.5 (close to 0), since most of the segregants should have the genotype 4, 5 or 6, which do not carry the SNP nucleotide (T in red), The superior segregants should carry the SNP nucleotide (genotype 1, 2 and 3), which results in the frequency of 0.5.

6.2.3.1 Segregation of homozygous SNP nucleotide in diploid segregants

In case of random segregation or when the SNP nucleotide is not important for the phenotype, the SNP frequency is expected to be around 0.5 (figure 6.2). However, when the SNP base is important for the phenotype, the segregation patterns deviate from 0.5 depending on the role the gene (mutation) is playing. For example, if the homozygous SNP in figure 6.2 (GG vs TT) is responsible for the phenotype, and if both alleles are required to render the good phenotype (recessive), then, only genotype 1 is expected in the superior pool (with SNP frequency of 1). However, in a quantitative trait, where more than one loci might be involved, this particular locus might be compensated by other genes in other loci and therefore, the frequency might slightly drop below 1, but still higher than 0.5. In the inferior pool (pool of segregants that do not show the phenotype), genotypes 2 to 6, with at most one of the SNP allele, might be inherited, and the frequency of SNP base G will be around 0.4 (4/10).

6.2.3.2 Segregation of heterozygous SNP nucleotide in diploid segregants

When a heterozygous SNP is responsible for the phenotype in the superior strain, the SNP frequency in both the superior and inferior pool is also different. For example, in figure 6.2 if the SNP base T (in red) is the responsible nucleotide variant for a dominant function (gain of function), then only the first three genotypes in figure 6.2 can be inherited in the superior pool, and the frequency of the SNP base T will be 0.5. However, since most of the SNPs in GS1.11-26 are homozygous, the influence of the heterozygous SNPs in the statistical analysis is minimal. Therefore, the frequency of the neighboring homozygous SNP G in figure 6.2, that also co-segregate with the heterozygous SNP T will have 66% inheritance (the first three genotypes in figure 6.2). In the inferior pool, only the last three genotypes that do not inherit the SNP base T from the superior parent can be represented in the pool. The neighboring homozygous SNP frequency in these three genotypes will then be about 33% (2Gs and 4Ts).

6.2.3.3 Effect of aneuploidy on SNP variant frequency

Another complexity of the genetic mapping with industrial strains is the presence of aneuploidy. When we analyze the whole genome sequence of the superior parent GS1.11-26 and the inferior parent Fseg25, we found that GS1.11-26 has three copies of chr IX and XVI, while Fseg25 carried three copies of chr III and chr X. As a result, the segregation of SNPs in these chromosomes is different from the one described above. In general, when the unique SNPs from the superior parent are used for the mapping (which is the case in our analysis), the presence of three copies of a chromosome in the superior parent results in an average SNP variant frequency above 0.5 during random segregation (because of over-representation of the chromosome from the superior parent). When a region is linked to the trait, the SNP variant frequency of the inferior pool drops below the average SNP frequency that is expected from random segregation (since most of the SNPs are inherited from the inferior parent). In this case, the SNP variant frequency of the superior pool will also be higher than the average. By the

same principle, the presence of three sets of chromosomes in the inferior parent results in an average SNP frequency below 0.5 in a random segregation (since the SNPs present in the superior parent are used for the analysis). Linked regions might therefore show SNP frequencies above the average, which might be closer to 0.5 or higher.

Because of these reasons, we decided to make the statistical inferences by comparing the SNP variant frequency of the superior pool and the inferior pool, rather than the deviation of the SNP variant frequency from the normal random variant frequency of 0.5. For that purpose, the log odds ratio of the SNP variant frequencies between the superior and inferior pool was calculated along with confidence intervals. A positive log odds ratio indicates a higher SNP variant frequency in the superior pool than in the inferior pool, and therefore linkage to the superior parent. A negative log odds ratio indicates a higher SNP variant frequency in the inferior pool than in the superior pool, and therefore linkage to the inferior parent. The p-value that corresponds to the statistically significant difference in the SNP frequencies among the two pools was then calculated based on the log odds ratio. An example of a plot is shown in figure 6.3. As can be seen in the top panel for chr IV, the average SNP variant frequency is around 0.5 in both the superior and inferior pools. The middle panel shows the log odds ratio which lies within the range of ±0.4, which corresponds to an SNP variant frequency that ranges between 0.4 and 0.6. This cutoff point was selected based on previous data, in which the average SNP variant frequency for random segregation oscillates between 0.4 and 0.6 (Swinnen *et al.,* 2012a). Another example is shown in figure 6.3 for chr XVI, where the SNP variant frequency is different from 0.5 in both the superior and inferior pool. As stated above, the superior strain GS1.11-26 has three copies of chr XVI and therefore the average SNP variant frequency was higher than 0.5 in both pools throughout that chromosome.

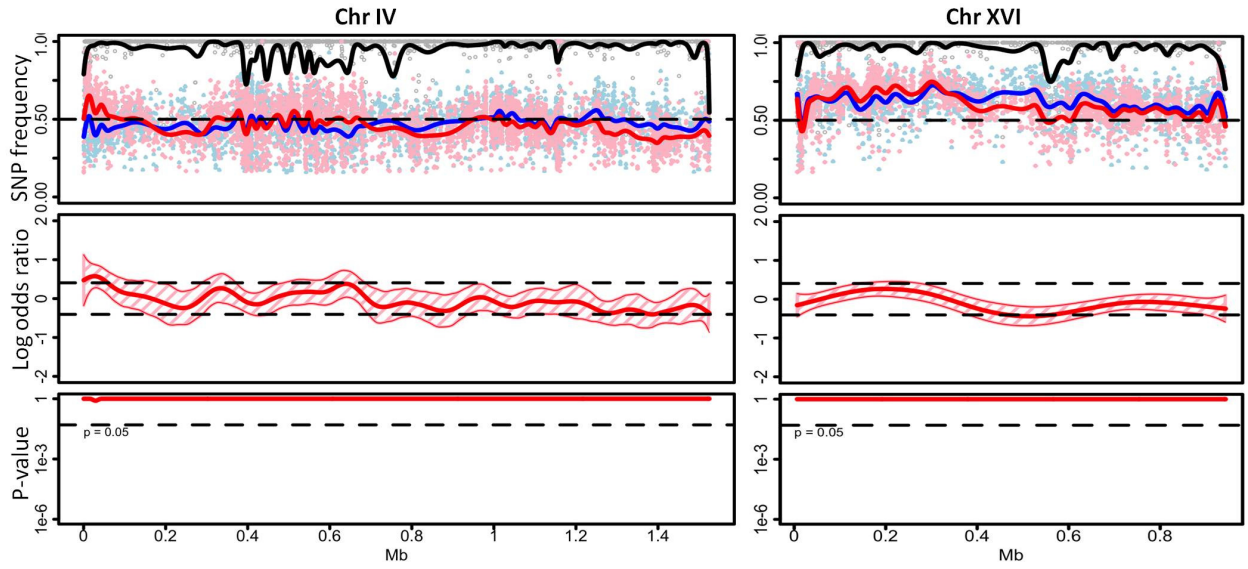


Figure 6.3. Comparison between SNP variant frequency of inferior pool and superior pool in chromosome IV and chr XVI. Top panel shows the SNP variant frequency of the inferior pool (blue dots) and superior pool (red dots) relative to the reference sequence. The blue and red lines represent the smoothened data for the inferior pool and superior pool, respectively. The SNP variant frequency and smoothened data of the superior parent GS1.11-26 are shown in grey circles and a black line, respectively and is used to estimate the baseline SNP variant frequency in the pools. Middle panel represents the log odds ratio (red line) along with the confidence interval (shaded regions) of the SNP variant frequency between the inferior and superior pool. Bottom panel represents the p-value for any difference between the SNP variant frequency of the inferior and superior pool (all values are close to 1, indicating no significant difference in the SNP variant frequency among the two pools.

6.3 Evaluation of loci that are linked to the phenotype

Using the above analysis, we have identified at least 3 QTLs. In this chapter we investigated the two QTLs (QTL1 on chr XV and QTL2 on chr XI) that showed the strongest linkage. Since the strain GS1.11-26 is derived from its parent strain HDY.GUF5, the only possible SNPs that can explain the fast xylose fermentation rate in GS1.11-26 should be SNPs that are not shared with HDY.GUF5 (in comparison with S288c). Therefore, when comparing the genome sequence of GS1.11-26 with that of HDY.GUF5, we focused on the SNPs that were different in the two strains. There were only a limited number of polymorphisms between the two strains in these two QTLs.

6.3.1 QTL1, chr XV

The strongest linkage was found in chr XV, where the xylose metabolism gene cassette has been integrated (Figure 6.4). We have shown in chapter 5 that part of the cassette that includes the *xylA* gene has been tandemly amplified in both alleles of the chromosomal locus in the superior parent GS1.11-26. The importance of multiple copies of the *xylA* gene for the high xylose fermentation performance was also verified (chapter 5). With this in mind, the significant deviation of the SNP frequency from 0.5 was expected since the amplified *xylA*-locus is essential

for the high xylose fermentation rate and should thus be present in all the segregants. However, since the calculated SNP variant frequency was lower than 1, there were apparently segregants present in the superior pool with only one of the two amplified *xylA*-alleles.

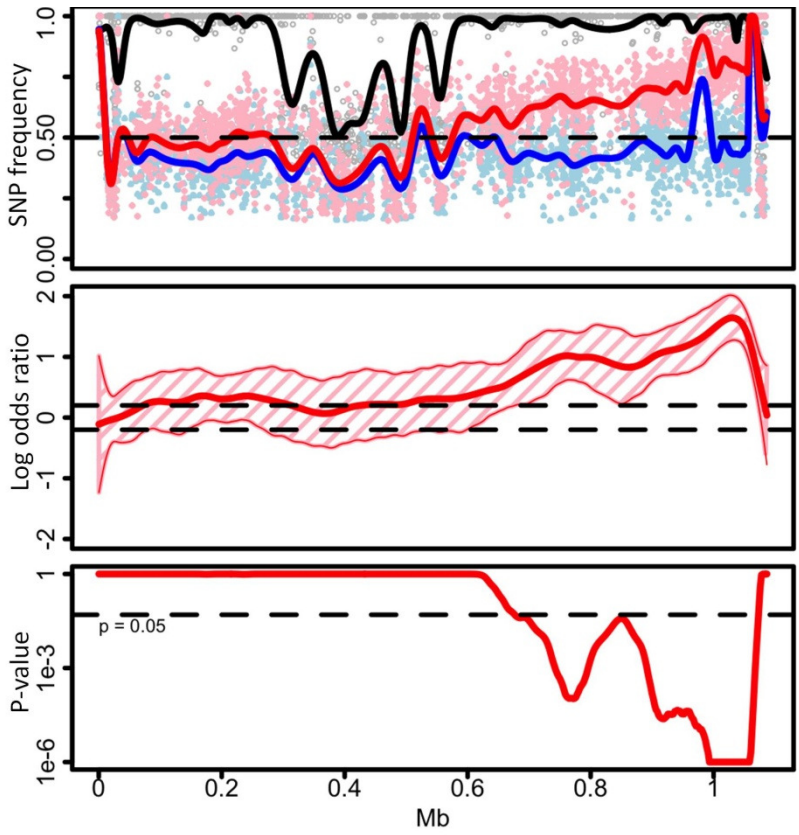


Figure 6.4. Comparison between SNP variant frequency of inferior pool and superior pool in chromosome XV. Top panel shows the SNP variant frequency of the inferior pool (blue dots) and superior pool (red dots). The blue and red lines represent the smoothening data for the inferior pool and superior pool, respectively. The internal SNP variant frequency and smoothened data of the superior diploid parent GS1.11-26 is shown in grey circles and black line, respectively. Middle panel represents the log odds ratio (red line) along with the confidence interval (shaded regions) of the SNP variant frequency between the inferior and superior pool. Bottom panel represents the p-value for the difference between the SNP variant frequency of the inferior and superior pool.

To estimate the number of alleles of the amplified *xylA*-locus in the pool, the sequence coverage at this locus in both the inferior and superior pool was compared relative to the coverage obtained for the superior parent GS1.11-26. The sequence data for GS1.11-26 showed a sequence coverage of about 8.8 fold higher than average at the *xylA*-locus (chapter 5). However, the sequence coverage of the superior pool was about 32% lower compared to the coverage for GS1.11-26 (6 fold higher than average compared to 8.8 fold) indicating that not all the segregants in the superior pool carried both alleles of the amplified *xylA*-locus. This means that about 68% of the segregants in the superior pool carried the amplified *xylA*-locus in both alleles, and the rest have it in only one of the alleles. On the other hand, the coverage for the

inferior pool at this locus was 50% lower compared to the coverage for GS1.11-26, suggesting that the individual segregants in the inferior pool carried only one of the alleles of the amplified *xylA*-locus. The presence of (at least) one amplified *xylA*-locus in the inferior segregants was expected, since all the segregants in the inferior pool were able to grow on D-xylose (but not ferment D-xylose) and should have at least one allele of the locus. Moreover, random segregation results in about 50 % chance of inheritance of the *xylA*-locus.

To experimentally determine the number of segregants in the superior pool that carried the amplified *xylA*-locus in both alleles, we evaluated all 27 superior segregants individually by PCR assays that specifically detect the *xylA* gene from GS1.11-26 and the *PYK2* gene from Fseg25 (Figure 6.5). Since the Xylose cassette has disrupted both alleles of the *PYK2* gene in GS1.11-26, a PCR reaction with primers upstream and downstream of *PYK2* can give a PCR product only on the allele from Fseg25, but not from GS1.11-26 (because the cassette is too large to be amplified by PCR). Another PCR assay using two primers inside the xylose cassette gives a PCR product only in the presence of the allele from GS1.11-26. Assuming a normal diploid genome without aneuploidy in the individual segregants, a positive PCR for only the GS1.11-26 allele was expected in the presence of the *xylA*-locus in both alleles. A positive PCR for the alleles of both Fseg25 and GS1.11-26 in a segregant was expected when one copy of the amplified *xylA*-locus is present and the other allele is inherited from Fseg25. With this assumption the PCR assays were performed in all 27 superior segregants. The superior and inferior parents were included as controls. As expected, all the 27 superior segregants showed a positive PCR for the GS1.11-26 allele, indicating that all the superior segregants carried at least one of the alleles of the amplified *xylA*-loci. However, 10 of the 27 segregants also showed a positive PCR result for the allele of Fseg25. This might indicate that these 10 segregants contain only one allele of the amplified *xylA*-locus while the other 17 (63%) inherited the amplified *xylA*-locus in both chromosomal alleles. This result was in close agreement with the result obtained by the sequence coverage analysis described above.

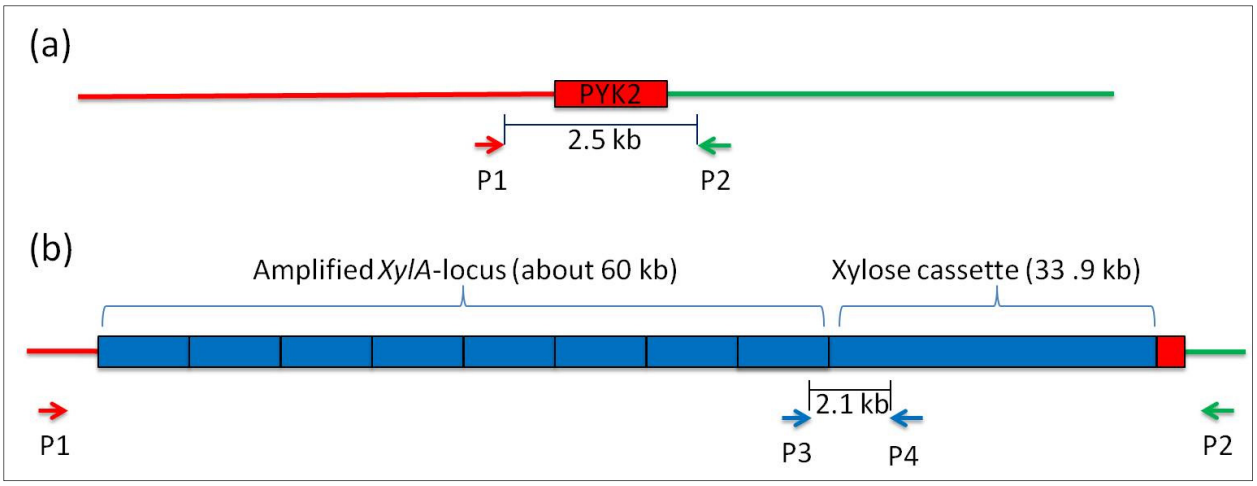


Figure 6.5. Schematic representation of the PCR strategy used to validate the presence of the amplified *xylA*-locus in the 27 superior D-xylose utilizing segregants isolated from the tetraploid parent GS1.11-26/Fseg25. (a) One of the two alleles from Fseg25; and (b) one of the two alleles from GS1.11-26 at the same *PYK2* locus, showing the amplified *xylA* region and the rest of the xylose cassette. The red bar at the right end of the xylose cassette stands for part of the *PYK2* sequence left at the region. P1 and P2 are the primers used to amplify the *PYK2* gene, which gives a positive PCR result in the strain Fseg25 and in segregants that carried at least one of the Fseg25 original *PYK2* alleles. In GS1.11-26, the PCR using primers P1 and P2 does not give a positive result because the region is too large to be amplified. P3 and P4 are primers that were used to specifically amplify a region from *xylA* gene present in GS1.11-26. The Fseg25 strain does not carry this gene, which therefore results in a negative PCR.

We then evaluated the fermentation performance of the 27 superior segregants to determine whether there was a correlation with the number of amplified *xylA*-alleles. However, there was no such correlation between D-xylose fermentation rate and the number of amplified *xylA*-alleles. Segregants carrying both alleles of the amplified *xylA*-locus did not always have a higher fermentation rate than those carrying only one amplified *xylA*-allele. The presence of rapidly xylose fermenting segregants with only one amplified *xylA*-locus probably indicates that one amplified *xylA*-locus is enough to support high xylose fermentation capacity, at least in the fermentation conditions used. This is in agreement with the result obtained for the GS1.4-14 strain, which carried only one amplified *xylA* allele but fermented xylose very well, with a rate close to that of the superior strain GS1.11-26 (chapter 2). Based on the results of the southern blot assay made with strain GS1.11-26, we could conclude that the number of copies of the *xylA* gene in both alleles was similar (about 8 copies of the *xylA* gene in each chromosomal locus) (chapter 5). Though a high copy number of *xylA* was clearly important for rapid xylose fermentation, doubling the number of *xylA* gene from 8 to about 16 did not double the xylose fermentation rate, but delivered only a moderate improvement. Precise determination of the copy number of the *xylA* gene required to obtain a maximum D-xylose fermentation rate might be important for the further development of industrial yeast strains with optimal performance in utilizing lignocellulosic feedstock hydrolysates.

Another QTL worth investigating is QTL1b on chr XV, close to QTL1a. From the SNP frequency data alone, it appears as if there is only a single QTL (Figure 6.4, top panel). However, the p-value analysis indicates two QTLs (Figure 6.4, bottom panel). In the region around 800kb, there are less SNPs compared to the other regions. This might be the reason for the separation of QTL1a and QTL1b in this region. However, this cannot rule out the possibility that a true separate QTL exists in the region and therefore, this region also needs further evaluation.

6.3.2 QTL2, chr XI

The second locus with a statistically significant linkage was located on chromosome XI (Figure 6.6). In this region, the average SNP variant frequency in the superior pool was about 0.6 while in the inferior pool was around 0.4. Though the SNP variant frequency in both pools was close to 0.5, the contrast between the two pools showed that the difference in the SNP frequency was statistically significant. As described above, a statistically significant deviation of the SNP variant frequency from 0.5 in the superior pool may not be as evident when evaluating diploid compared to haploid segregants. Instead, a significant difference between the SNP variant frequency of the superior and inferior pools is a better parameter to identify QTLs in diploid segregants, especially when a gain of function mutation is involved in the phenotype (see section 6.1.3, Segregation in diploid segregants (tetraploid parent)). Using this approach, the first 200kb of the chromosome showed the strongest linkage.

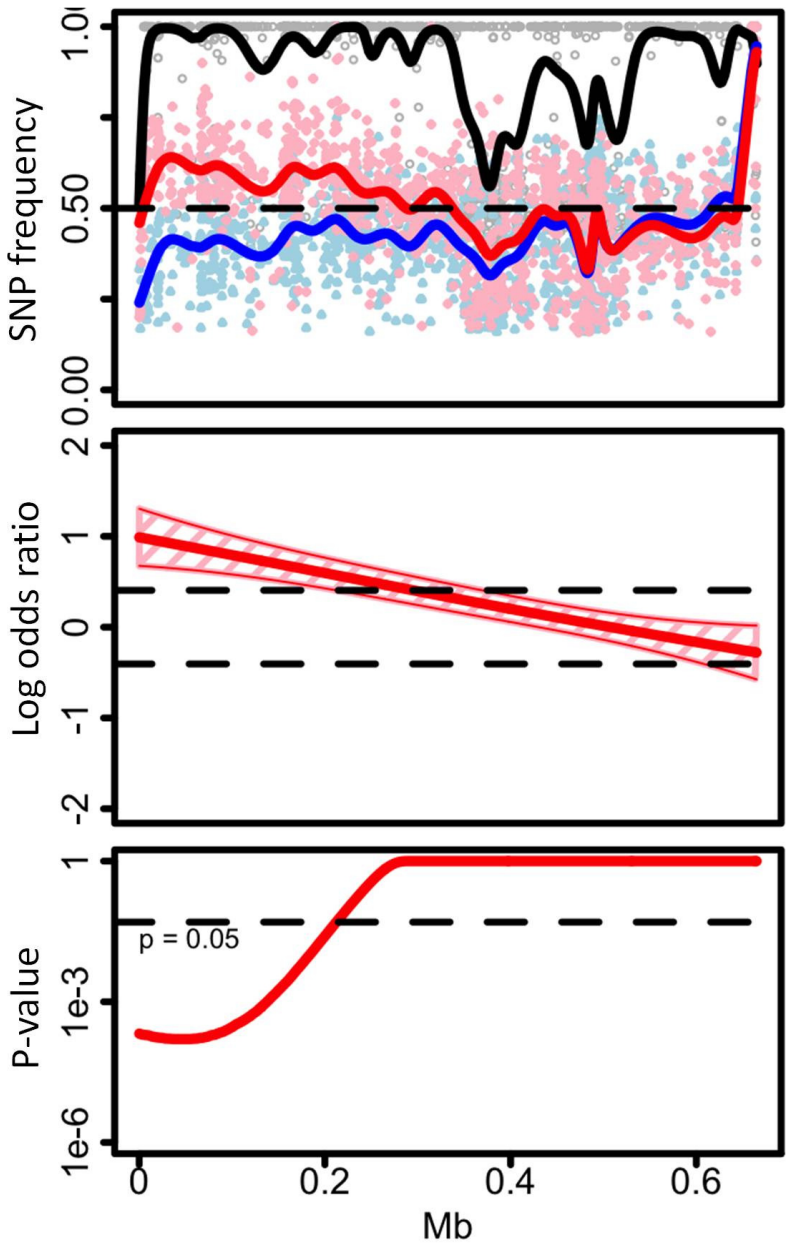


Figure 6.6. Comparison between SNP variant frequency of inferior pool and superior pool for chromosome XI. Top panel shows the SNP variant frequency of the inferior pool (blue dots) and superior pool (red dots). The blue and red lines represent the smoothening data for the inferior pool and superior pool, respectively. The internal SNP variant frequency and smoothened data of the superior diploid parent GS1.11-26 is shown in grey circles and black line, respectively. Middle panel represents the log odds ratio (red line) along with the confidence interval (shaded regions) of the SNP variant frequency between the inferior and superior pool. Bottom panel represents the p-value corresponding to the difference between the SNP variant frequency of the inferior and superior pool.

When the polymorphisms between the parent HDY.GUF5 and the evolved strain GS1.11-26 in this locus were compared, there were only 12 SNPs that were either located within an ORF or close to an ORF (500bp upstream and 300bp downstream) (Table 6.2). To further narrow down

the region, we performed fine mapping of the locus with allele specific PCR assay in the 27 individual segregants. Since all the segregants have a diploid genome (therefore, there are two possible alleles), the scoring method to determine the SNP variant frequency for individual SNPs had to be modified. We first searched for homozygous SNPs that are unique to either the superior diploid parent GS1.11-26 or the inferior diploid parent Fseg25. We then performed an allele specific PCR assay to detect only these SNPs in the individual segregants. When an individual segregant carried the SNP nucleotide from both parents, it was given a score of 1, (since it carries the SNP nucleotide only once). When it carried the SNP nucleotide from only one of the two parents (carries two times the same nucleotide), it was given a score of 2. The 27 segregants were evaluated in this way and the results allowed in reducing the strongly linked area to about 50 kb (Figure 6.7).

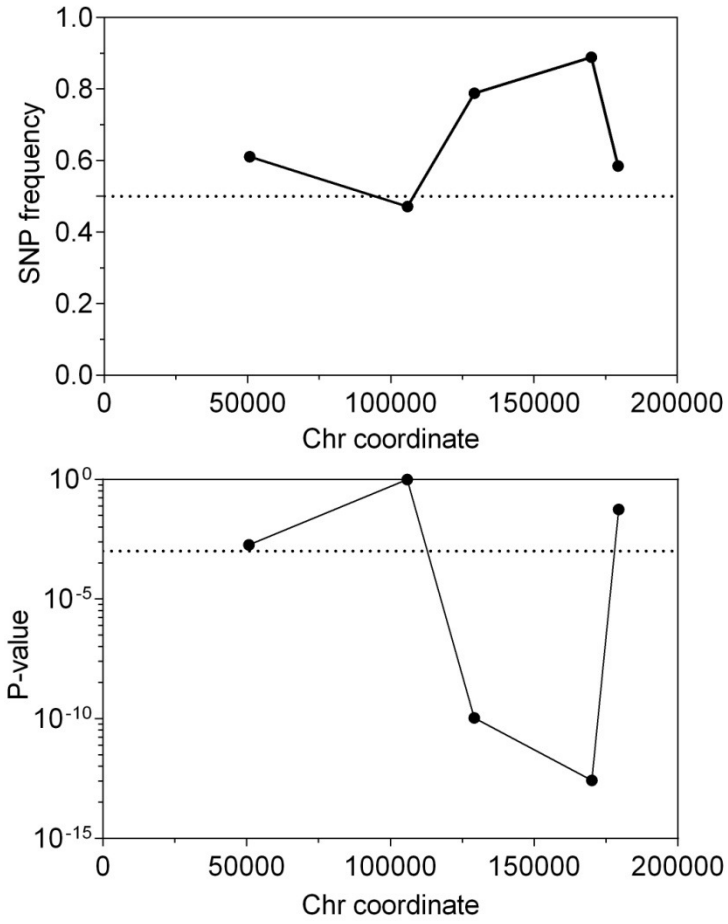


Figure 6.7. SNP variant frequency (top) and p-value (bottom) for the 5 selected SNPs in the individual segregants of the superior pool at QTL2. The p-value was calculated based on the deviation of the SNP frequency from 0.5.

This region contains about 24 genes, but only 5 genes had mutations unique to GS1.11-26 compared to its parent strain HDY.GUF5 (table 6.2). One of the mutations is located upstream of the genes *SDH1* (Crick strand) and *AVT3* (Watson strand) (which might be in the promoter of

both genes), and another one is a synonymous mutation in the gene *MCR1*. The other 3 are missense mutations in *NNK1*, *ELF1* and *SDH3*.

Table 6.2. SNPs found in GS1.11-26 relative to the wild type HDYGUF5 in the QTL2 locus. Only SNPs in an ORF or 500bp upstream or 300bp downstream of an ORF are shown. SNPs that alter the amino acid sequence are shown in bold. All the SNPs were heterozygous in the evolved strain. The function of the gene products was taken from Saccharomyces Genome Database (http://www.yeastgenome.org/)

Chr coordinate	Reference base	SNP base	Coverage	Type of mutation	Gene	Function
22034	C	T	18	Upstream	*JEN1*	Monocarboxylate/proton symporter of the plasma membrane
41069	C	T	30	**Missense**	**UBA1**	Ubiquitin activating enzyme (E1)
54146	G	A	33	**Missense**	**EAP1**	eIF4E-associated protein,
106249	G	A	43	**Missense**	**FAS1**	Beta subunit of fatty acid synthetase
119732	C	T	48	**Missense**	**ZRT3**	Vacuolar membrane zinc transporter
119863	C	T	40	Synonymous	ZRT3	Vacuolar membrane zinc transporter
121834	C	T	46	**Missense**	**TPO5**	Protein involved in excretion of putrescine and spermidine
129894	G	A	21	**Missense**	**NNK1:**	Protein kinase; implicated in proteasome function
153491	G	A	36	**Missense**	**ELF1**	Transcription elongation factor that contains a conserved zinc finger domain
167338	C	T	18	Synonymous	*MCR1*	Mitochondrial NADH-cytochrome b5 reductase, involved in ergosterol biosynthesis
171522	C	A	32	Upstream	*SDH1/ AVT3*	SDH1=Flavoprotein subunit of succinate Dehydrogenase : AVT3=Amino acid Vacuolar Transport
179847	G	A	42	**Missense**	**SDH3**	Subunit of both succinate dehydrogenase and of TIM22 translocase

All 5 mutations were heterozygous in the superior parent GS1.11-26 compared to the inferior parent Fse25 (which has the same genome sequence as the reference genome from S288c). We then evaluated the possible involvement of all 5 mutations by deleting either the mutant or the wild type allele in the GS1.11-26 strain. First, the ORF of *SDH3, MCR1, ELF1* and *NNK1*, as well as the sequence between *AVT3* and *SDH1* (containing the SNP) were replaced by a kanamycin resistance marker. This resulted in the deletion of either the mutant or the wild type allele (Figure 6.8). As a result, deletion strains that possess only the wild type allele or the mutant allele were obtained. These strains were identical with respect to their genome, except for the particular allele under study; therefore, comparison of the two strains allows to evaluatethe effect of either allele. Each couple of strains with reciprocally deleted alleles was then evaluated for fermentation performance in D-xylose medium.

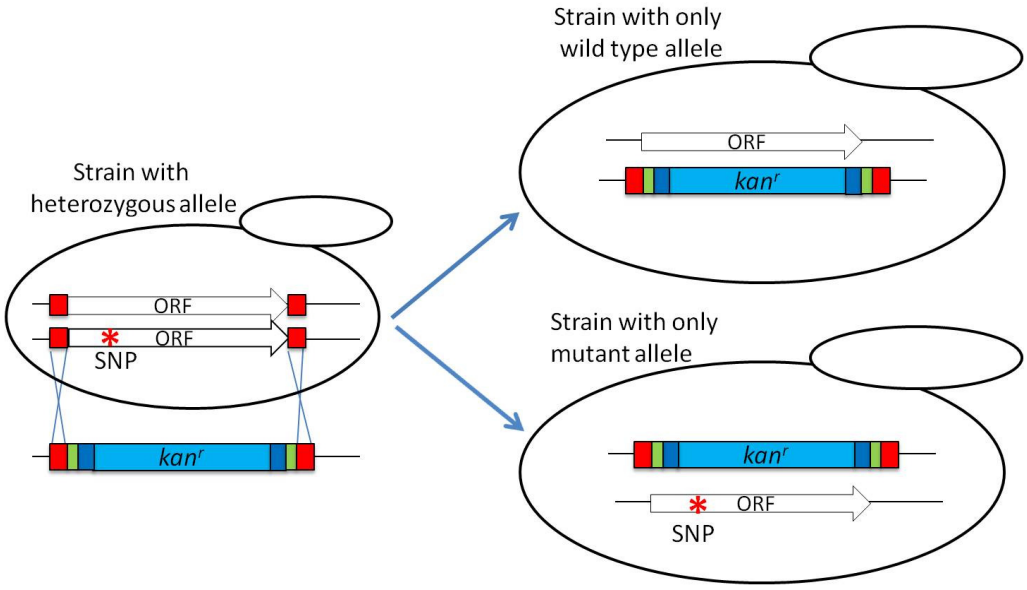


Figure 6.8. Strategy used for construction of the reciprocal deletion strains with a different allele for RHA using the ΦC31 integration system. A strain with two heterozygous alleles was transformed with a PCR product containing a kanamycin resistance marker flanked with a homologous region (red) on either side of the ORF to be deleted. A few transformants were picked up and the region with the mutation in the gene under investigation was sequenced. Transformants with a single gene deletion and containing either the wild type allele or the mutant allele of the gene were selected for evaluation. Bars in dark blue represent the attP (left) and attB (right) sequences that are used to remove the marker using the ΦC31 integrase system. The green bars represent adaptors for PCR amplification of the marker.

As can be seen from figure 6.9 reciprocal deletion of the three genes (*SDH3, MCR1 and ElF1*) and the region between *AVT3* and *SDH1* did not result in any clear difference in the rate of D-xylose fermentation compared to the original GS1.11-26 strain. However, a very clear difference in D-xylose fermentation rate was observed among the strains that are reciprocally deleted for the *NNK1* allele. The strain that carried only the wild type allele of *NNK1* showed a much slower rate of fermentation compared to the strain that carried only the mutant allele (*nnk1*^{S807N}). The fermentation rate by strains carrying only the *nnk1*^{S807N} allele of the superior parent GS1.11-26 was also slightly better than that of the original GS1.11-26 strain (that carried the two different alleles).

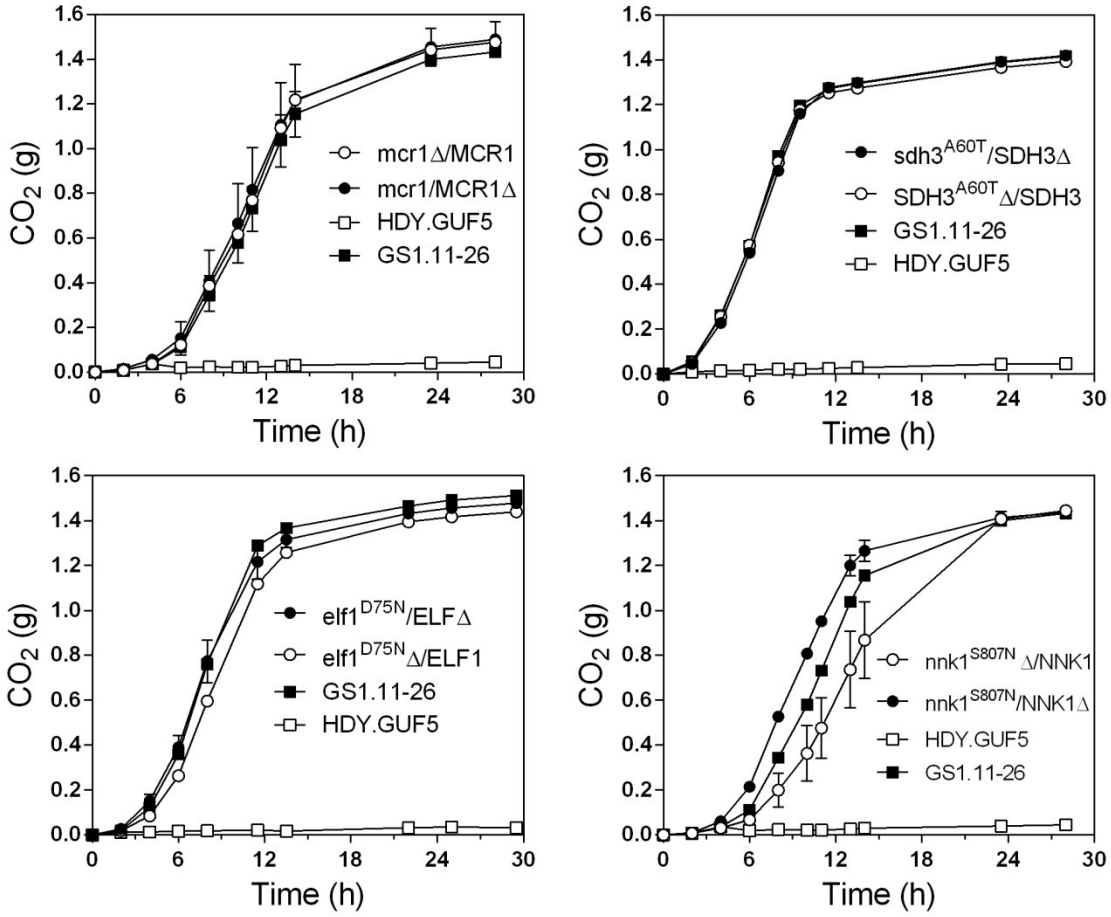


Figure 6.9. D-xylose fermentation profile of strains that are reciprocally deleted for the four ORFs in the GS1.11-26 strain background. At least two independent transformants were evaluated for each gene except for ELF1, for which two transformants with a mutant allele and one transformant with a wild type allele were evaluated. Genes with the mutant allele are indicated with small letter with the amino acid change, while genes with the wild type allele are shown in capital letter. *MCR1* had a synonymous mutation. GS1.11-26 carried the heterozygous alleles for all the genes tested.

6.3.3 Other putative QTLs

The QTL3, located on chr X, includes a region of about 150kb (Figure 6.10). From the whole genome sequence analysis of the two parent strains described above, we know that the inferior parent has three copies of chr X. Therefore, the SNP variant frequency in this region is expected to be lower than 0.5 in the case of random segregation. This is indeed what we found. Both the inferior and the superior pools showed an SNP variant frequency of about 35% in the majority of the locations in chr X. However, a statistically significant deviation of the SNP variant frequency of the two pools from each other was observed at QTL3, due to the higher SNP variant frequency in the superior pool at this location compared to the average in the rest of the chromosome. Sequence comparison of the evolved strain GS1.11-26 and its parent HDY.GUF5 showed several polymorphisms. Further investigation is required to link one or more

of the polymorphisms between the two strains in this QTL3 with the efficient D-xylose fermentation rate.

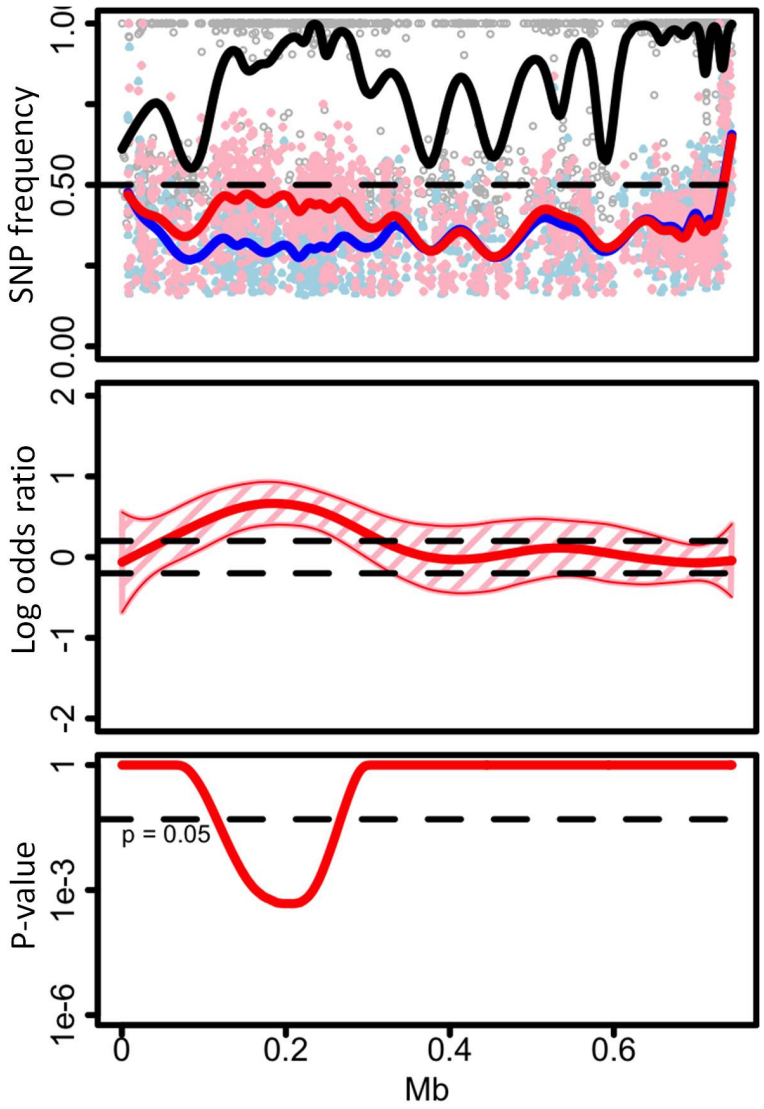


Figure 6.10. Comparison between SNP variant frequency of inferior pool and superior pool in chromosome X. Top panel shows the SNP variant frequency of the inferior pool (blue dots) and superior pool (red dots). The blue and red lines represent the smoothening data for the inferior pool and superior pool, respectively. The SNP variant frequency and smoothened data of the superior parent GS1.11-26 is shown in grey circles and black line, respectively. Middle panel represents the log odds ratio (red line) along with the confidence interval (shaded regions) of the SNP variant frequency between the inferior and superior pool. Bottom panel represents the p-value corresponding with the difference between the SNP variant frequency of the inferior and superior pool.

In addition to the three mentioned QTLs, analysis with less stringency for the difference in the odds ratios between the two pools indicated three other loci that might be weakly linked to the efficient xylose fermentation phenotype. In chr III, the SNP variant frequency in both pools

dropped below 0.5 (appendix 2) due to the presence of three copies of this chromosome in the inferior parent Fseg25. However, there was statistically significant deviation of the SNP variant frequency between the two pools, for the same reason described above for QTL3 in chr X.

6.4 Discussion

In this chapter, the genetic basis behind the rapid xylose fermentation capacity in strain GS1.11-26 has been investigated using pooled-segregant whole genome sequence analysis. In the conventional PSS, haploid parent strains are always used as starting strains. However, many industrial or sometimes also natural yeast strains are diploid, polyploid or aneuploid, and it is extremely difficult to obtain a pure haploid descendant (i.e. with a single copy of all chromosomes) from such strains that also shows the same quantitative trait of interest as its parent. These phenomena strongly hamper the use of the method for the genetic mapping of industrially important traits in such strains.

To circumvent this limitation, we have investigated the possibility of using PSS to map QTLs in a diploid strain without the need for isolation of haploid descendants. We have investigated the genetic basis responsible for the rapid D-xylose fermentation trait in the industrial diploid strain GS1.11-26 by crossing it with another diploid strain Fseg25. Both parent strains contain 3 copies of two of their chromosomes (which therefore, could be called aneuploids), which made the mapping even more complicated.

The strain GS1.11-26 carried the xylose/arabinose metabolism cassette in both alleles at the *PYK2* locus in chromosome XV. In addition, the *xylA* gene has been amplified to about 8 copies in both alleles of the chromosomal locus and this is required for the high xylose fermentation capacity. Therefore, this locus served as a positive control for the evaluation of the technique, since all superior segregants should carry the cassette, and as a result, a linkage to this locus is expected to occur if the method works properly. We compared the SNP variant frequency in the two pools of segregants; a superior pool that contains 27 segregants with rapid xylose fermentation ability, and an inferior pool that contains 27 segregants that were not able to ferment xylose. As anticipated, a statistically significant linkage was found at the locus where the xylose/arabinose cassette had been integrated. We evaluated the individual segregants for the presence of the cassette and found that all the segregants carried the cassette in at least one of the alleles, while 63% carried the cassette in both alleles. This confirms the importance of the locus for the rapid D-xylose fermentation rate as well as the validity of the method we applied.

Using the statistical tool that was used for the conventional PSS, no other loci could be identified, as the SNP variant frequency did not show statistically significant deviations from 0.5. Next, we made predictions of the possible meiotic segregation pattern of polymorphisms in diploid meiotic segregants that can be obtained from a tetraploid parent. Accordingly, random segregation of chromosomes does not follow the regular 2:2 segregation seen in haploid segregants. Therefore, we used a modified statistical analysis tool to evaluate the difference in

SNP variant frequency between the two pools. As a result, we were able to identify at least two other QTLs with statistically significant linkage.

The absence of other loci with a linkage as strong as QTL1 might be due to the capacity of more than one locus in supporting a single function in sustaining high xylose fermentation capacity. As described above, QTL1 contains the heterologous *xylA* gene, which is essential for the phenotype. Since this gene cannot be complemented by other genes in other loci, all segregants in the superior pool should inherit the gene; therefore, the SNP variant frequency in QTL1 is close to 1. If more than one locus is able to support the same function, the segregant can inherit the allele from either of the loci, and therefore the SNP variant frequency for those loci will drop below 1. Another reason for the SNP variant frequency to drop to a lower value than expected is the presence of heterozygous alleles in the superior diploid parent important for the phenotype. As indicated in figure 6.2, the highest possible SNP variant frequency in the superior segregants in this case is 0.5. In the inferior pool, the SNP variant frequency in this case drops below 0.5 (figure 6.2). Nevertheless, we were able to identify the QTLs using the modified statistical analysis tool.

QTL mapping with a pool of diploid segregants has been performed recently (Parts *et al.,* 2011). In that report, two phenotypically distinct strains were crossed for many generations, after which, the haploid and diploid pool of 'segregants' were grown in selective conditions (high temperature) to enrich for the beneficial alleles. The total DNA from the pools was finally sequenced. Though the study demonstrated that it is possible to carry out genetic mapping with a pool of diploid strains, it originally started with haploid parent strains. In that case, the analysis might not be significantly different from the regular genetic mapping since the SNP variant frequency cannot be influenced by heterozygous alleles in the original parent strain since they used haploid parents. In our study, however, a diploid strain was used as the superior parent. The fact that we could demonstrate that QTL mapping can be performed with pools of diploid segregants originating from polyploid/aneuploid parents will significantly strengthen the applicability of this genetic mapping strategy. It will allow performing efficient genetic mapping of industrially important traits without losing the genetic architecture of the original strains, which is often complex in industrial strains or natural isolates with an interesting trait.

We have identified a mutation in the gene *NNK1* in QTL2 that is involved in high D-xylose fermentation rate. Though we have not yet confirmed the usefulness of the mutation in *NNK1* in the original parent strain, the RHA analysis showed that the mutant allele *nnk1*^{S807N} is important for the high D-xylose fermentation rate in the evolved strain. *NNK1* encodes a protein kinase that was previously implicated in proteasome function and also found to be linked to nitrogen metabolism through the association with the TORC1 subunits and the

NAD(+)-dependent glutamate dehydrogenase, encoded by *GDH2* (Breitkreutz *et al.,* 2010; Cagney *et al.,* 2001). Nnk1 phosphorylates Gdh2, which is a means of regulation of Gdh2 activity. Higher D-xylose to ethanol conversion rate has been previously linked to modulation of nitrogen metabolism by deleting *GDH1* and overexpression of *GDH2* in strains expressing the XR/XDH pathway (Roca *et al.,* 2003). This was performed to obtain an adjustment of the redox imbalance that was created by introduction of the XR/XDH pathway. Since our strain was constructed using the XI pathway, it is not expected to be affected in the redox balance during conversion of xylose to xylulose. The mechanism of action of the mutation (*nnk1*^{S807N}) in sustaining rapid D-xylose fermentation in our strain is not known and requires further investigation.

General conclusions and recommendations

Conclusions

Bioethanol production from non-food biomass such as energy crops, agricultural wastes and forest residues as substrates is nowadays advancing towards commercial production scale. Bioethanol produced from such lignocellulosic biomass is known to have a high potential for the overall replacement of fossil fuels used for the transportation sector, and for establishing in this way a significant reduction in GHG emissions. However, bioethanol production from such biomass is not yet economically viable. One of the biggest hurdles for economically competitive ethanol production from lignocellulosic biomass is the lack of microorganisms that efficiently convert all the sugars present in the lignocellulose hydrolysates into ethanol under industrial conditions. Currently, the baker's yeast *Saccharomyces cerevisiae* remains the best organism for industrial ethanol production because of several reasons, including its high rate of fermentation of hexose sugars and its high tolerance to ethanol and to inhibitors in the lignocellulosic hydrolysates. Unfortunately, *S. cerevisiae* is unable to metabolize pentose sugars, particularly D-xylose, which represents up to 35% of total sugars in second generation feedstocks. Thus, development of a D-xylose-utilizing, highly-robust industrial strain of *S. cerevisiae* is crucial for cost effective and sustainable production of ethanol from lignocellulosic material.

The main objective of this study was the construction of a robust D-xylose utilizing industrial *S. cerevisiae* strain. The study started with one of the most widely used industrial bioethanol production strains, Ethanol Red. All the genes known to be required for D-xylose and L-arabinose utilization had been integrated into the genome of Ethanol Red (by the group of Prof E. Boles, University of Frankfurt, Germany). The key gene for the xylose utilization pathway was *xylA*, coding for xylose isomerase, which originated from the bacterium *C. phytofermentans.* In spite of efforts to optimize the enzyme through codon usage adaptation, its total activity remained too low in the original recombinant strain HDY.GUF5. As a result, the recombinant strain was not able to ferment D-xylose.

In this study, a combination of random mutagenesis, genome shuffling and evolutionary adaptation in D-xylose medium was employed to improve the D-xylose utilization rate of the recombinant strain. Consequently, the final strain GS1.11-26, showed the highest D-xylose to ethanol conversion yield compared to any other engineered *S. cerevisiae* strains reported so far. The work also demonstrated that combining a rational metabolic engineering strategy with a systematic approach of random strain modification (mutagenesis, genome shuffling and evolutionary adaptation) provides an efficient tool for development of strains with high D-

xylose fermentation capacity. The method might also be applied for improvement of other industrial properties, like stress tolerance. Though the contribution of the genome shuffling step was not confirmed, the mutagenesis and evolutionary adaptation steps were shown to clearly contribute to the improvement of the strain during the strain engineering strategy. The best D-xylose utilizing strain GS1.11-26 was shown to perform very well in three lignocellulosic hydrolysates (Arundo donax, wheat straw/hay mixture and spruce hydrolysates). However, major challenges arose when the strain also showed a reduction in aerobic growth rate, ethanol accumulation capacity and tolerance to acetic acid. Such trade-off situations are commonly observed constraints of random strain modification strategies (Sonderegger and Sauer, 2003; Wenger *et al.,* 2011).

To solve these problems, we used a meiotic recombination technique that recombined the genome of the GS1.11-26 strain with that of a segregant of Ethanol Red or a robust inhibitor tolerant strain. Using this method, three new hybrid strains (GSE16, GSF335 and GSF767) were constructed. These strains demonstrated efficient D-xylose utilization, higher inhibitor tolerance and high aerobic growth rates. As a result, they have a great potential for direct application in industrial bioethanol production. However, there is still much room for improvement of the D-xylose utilization rates in the three strains, as they lost some of the xylose fermentation capacity in comparison to the evolved strain GS1.11-26.

One strategy to further improve the high D-xylose fermentation rate while maintaining the industrial robustness of the strain is to identify the genetic determinants of the high D-xylose utilization rate in GS1.11-26, and to transfer the relevant mutant genes into a clean industrial strain that was not exposed to the mutagenesis. As the performance of industrial strains varies in different industrial conditions, a strain that outperforms in one industrial condition may not show the same superiority in another industrial condition. As a result the yeast fermentation industries use different yeast strains that are best suited in their specific industrial conditions. Hence, when the genetic basis behind the efficient D-xylose utilization in GS1.11-26 has been identified, this genetic information can be transferred to various target industrial strains, suitable for different applications. This will not be limited to bioethanol production strains but can be expanded to strains that are used for production of other biofuels, like isobutanol, or bio-based chemicals like specialty lipids and organic acids, using lignocellulose biomass as a substrate.

With the perspective of unraveling the genetic basis behind the high D-xylose utilization capacity in GS1.11-26, its genetic differences from the parent strain HDY.GUF5 have been determined by comparing their genome sequence. By doing so, it was found that the *xylA* gene has been amplified about 8 to 9 fold in the evolved strain. The contribution of this genetic

change as major and essential element for the high D-xylose utilization rate was also confirmed. The high copy number of *xylA* was associated with high XI activity, which is essential for rapid D-xylose utilization. However, over-expression of the XI in the original strain did not render any improvement in D-xylose utilization rate. This clearly demonstrated that in addition to the high XI activity other genetic changes in the genome of GS1.11-26 were important for the high D-xylose fermentation capacity.

To identify some of the remaining genetic changes important for the high D-xylose utilization rate in GS1.11-26, a QTL mapping technique was performed using pooled-segregant whole genome sequence analysis. Unlike the regular PSS that starts with crossing of two haploid strains, two mating competent diploid strains were used as starting parents. With a modified statistical analysis tool, it was possible to find statistically significant linkage in at least three QTLs, one of which is the expected locus where the heterologous *xylA* gene had been integrated. In addition, the gene *NNK1* (coding for a protein kinase) has been identified in the second QTL as a promising candidate gene for improvement of D-xylose utilization in this strain background. How *NNK1* can be linked to D-xylose fermentation in this strain background and its precise contribution to the improvement of D-xylose fermentation needs further study. In any case, the mutation in *NNK1* in the evolved strains seems to have a positive role with respect to D-xylose fermentation. Moreover, since the D-xylose fermentation rate by the GS1.11-26 strain that carried only the mutant *NNK1* allele was even higher than that of the original GS1.11-26 strain that carried both the mutant and the wild type alleles, replacing the wild type allele by a second copy of the mutant allele might further improve the D-xylose fermentation rate.

Identification of QTLs responsible for a trait-of-interest is an important but up to now also the most difficult step in inverse metabolic engineering. To date, QTL mapping in yeast has been limited to haploid strains. This strongly hampered the application of QTL mapping in industrial strains or natural isolates, which are mostly diploids, sometimes polyploid or aneuploid. This study revealed for the first time that QTL mapping can be performed by crossing mating competent diploid and/or aneuploid yeast strains without the need to isolate haploid derivatives. Though the mating competent diploid strain GS1.11-26 used in this study was generated spontaneously by mutagenesis, polyploid strains can easily be made mating competent by switching their mating type (Pavelka *et al.,* 2010). Therefore, the method can be applied to identification of genetic determinants for various industrial traits like aroma profile in wine and beer strains, stress tolerance in strains that are used in biofuel production (both first and second generation) and production of other bio-based chemicals.

In the present work, not all the genes responsible for the improved trait in GS1.11-26 have been identified. However, the heterologous gene *xylA* integrated in chr XV served as a marker

to test the validity of the QTL mapping. The strong linkage at this locus confirmed the potential applicability of the method in strains with diploid, polyploid and/or aneuploid genome. Therefore, this study has laid a strong foundation for more extensive QTL mapping of traits-of-interest in industrial strains and natural isolates with complex genetic architecture as well as strains generated by artificial mutagenesis. This appears to have a very promising outcome in dealing with the identification of complex traits in a more natural genetic setting, which will clearly constitute a step forward for inverse metabolic engineering.

Recommendations

GS1.11-26 remains the best strain for D-xylose utilization developed in this study. The negative mutations, particularly the slow aerobic growth rate, hampered the possible use of the strain directly in industrial bioethanol plants. Curing this strain by meiotic recombination with a haploid segregant of ER or with the most inhibitor tolerant strain isolated from our strain collection improved all the negative properties that were identified, but the D-xylose utilization rate by the cured strains was reduced compared to GS1.11-26. One reason may be that there is a causal link between high xylose utilization rate and slow aerobic growth rate. In that case, it will be very challenging to obtain a strain with high growth rate and fast xylose utilization capacity. One would have to make a compromise between the two traits. An alternative cause might be that the beneficial mutations and the negative mutations reducing the aerobic growth rate are functionally independent but located very close to each other in the genome; as a result, both types of mutations are inherited together in the F1 generation strains. This problem could be alleviated through extensive meiotic recombination using multiple inbreeding in order to increase the number of crossing-overs in the genome. In the present study, the meiotic recombination was done only once, and in the case of crossing with ER17, only about 100 segregants were screened. Increasing the number of inbreeding steps and the number of segregants screened might increase the chances of finding a robust strain with a minimum number of negative mutations while maintaining the high xylose fermentation capacity. Moreover, using the diploid Ethanol Red strain, instead of its haploid derivative as a mating partner for GS1.11-26, might also be beneficial, because the segregants would be able to retain more of the beneficial properties of Ethanol Red.

An alternative approach of constructing a more robust strain would be to implement an inverse metabolic engineering strategy and was already started in this study. As described above, the most important genetic element (amplification of *xylA*) has been identified in the current work. However, additional genetic factors are also involved, without which the high D-xylose fermentation rate cannot be attained. QTLs other than QTL1 and QTL2 have not yet been investigated in this work. Studying the remaining QTLs might reveal additional important mutations. In addition, QTL1 itself appears to have two independent regions. If confirmed, it

will be necessary to evaluate all the genes that show genetic differences among the evolved GS1.11-26 strain and the parent strain HDY.GUF5 located in this double QTL. This will also rule out the possibility that there are important mutations located close to the *xylA*-locus. Identification of the remaining causative genetic elements and transferring all of them into a fresh Ethanol Red strain might be a good alternative strategy. This would allow to build a new superior strain and lose all negative mutations and also the additonal background mutations, which actually may have difficult to identify negative effects in industrial scale. In addition, the beneficial mutations can also be transferred to the three new cured strains derived from GS1.11-26 (if they do not yet harbor these beneficial mutations) since these strains show more robustness in some conditions relevant for fermentation of lignocellulose hydrolysates than the original industrial strain Ethanol Red.

Materials and Methods

Strains and growth conditions

The *S. cerevisiae* strains utilized in this study are listed in Table 1. Yeast cells were propagated in yeast extract peptone (YP) medium (10 g/L yeast extract, 20 g/L bacteriological peptone) supplemented with either 20g/L D-xylose (YPX) or 20g/L D-glucose (YPD). For solid plates, 15g/L Bacto agar was added. For batch fermentation, either YP medium or synthetic complete medium (1.7 g/L Difco yeast nitrogen base without amino acid and without ammonium sulfate, 5 g/L ammonium sulfate, 740 mg/L CSM-Trp and 100 mg/L L-tryptophan) supplemented with D-xylose or D-glucose/D-xylose mixture was used. For selection of strains expressing the *Kan*MX resistance marker, 200 mg/L geneticin was added to the medium. Yeast strains were maintained at -80 °C in stock medium composed of YP and 26% glycerol.

Table 7.1 *S. cerevisiae* strains used in the study.

Yeast strain	Main characteristics	Source/reference
Ethanol Red	Industrial bioethanol production strain, *MAT**a**/α*	Fermentis, a division of S. I. Lesaffre, Lille, France
HDY.GUF5	Ethanol Red**;** *pyk2::xylA*; *XKS1; TAL1; TKL1; RPE1; RKI1; HXT7;AraT; AraA; AraB;AraD; TAL2; TKL2*	Goethe University of Frankfurt, Germany
M315	HDY.GUF5 + 3 h mutagenesis in 3% EMS, *MATα/α*	This study
M492	HDY.GUF5 + 4 h mutagenesis in 3% EMS, *MAT**a**/α*	This study
GS1.11-26	HDY.GUF5**,** M315 and M492 + genome shuffling and evolutionary adaptation, *MATα/α*	This study
TMB3400	USM21 HIS3::YIpXR/XDH/XK + mutagenesis and selection	(Wahlbom *et al.,* 2003)
GSE16	Hybrid strain obtained by meiotic recombination of GS1.11-26 and Ethanol Red segregant ER17	This study
GSF335	Hybrid strain obtained by meiotic recombination of GS1.11-26 and Fseg25	This study
GSF767	Hybrid strain obtained by meiotic recombination of GS1.11-26 and Fseg25	This study
JT21653b	Baker's yeast *S. cerevisiae*, obtained after screening a collection of yeast strains for high inhibitor tolerance in spruce hydrolysate, *MAT**a**/α,* triploid.	MCB, KU Leuven
Fseg25	Segregant of JT21653b selected for high inhibitor tolerance in spruce hydrolysate fermentation, *MAT**a**/**a***	This study
ER17	Segregant of Ethanol Red, exhibiting high acetic acid tolerance, *MAT**a***	MCB, KU Leuven
MV1000	Mating type tester strain, *MAT**a***, bar1Δ	MCB KU Leuven
MV1000	Mating type tester strain, *MATα, sst2*Δ	MCB KU Leuven

Mutagenesis

Overnight-grown yeast cells were harvested, washed twice with Sodium phosphate buffer (pH 7), and re-suspended in 1mL sodium phosphate buffer at a cell concentration of 2 x 10^8 cells/mL. Five different samples were treated with a final concentration of 3% Ethyl Methanesulfonate (EMS) or only phosphate buffer (as control) for different time intervals at 30°C. The EMS was subsequently neutralized by washing twice with freshly prepared 5% sodium thiosulphate. The cell pellets were then re-suspended in sterile 500 µL milliQ water and plated in aliquots of 100 µL onto both YPX and YPD plates. To estimate the percentage survival after mutagenesis, colonies of EMS treated cells from the YPD plates were counted and the ratio relative to that of untreated cells was calculated.

Genome shuffling

For genome shuffling, cells selected based on growth on D-xylose and sporulation efficiency were sporulated in 1% potassium acetate medium. After 7 days at 23°C, asci were harvested and spores were purified (Hou, 2009). The purified spores from each strain were mixed together and allowed to germinate for 2 h in YPD medium. Exponentially growing cells from a MATα/α diploid strain (M315) were mixed with the germinated spores. The cells were allowed to mate in 40 mL YPD in a shaking incubator at 70rpm for 48 h. To select D-xylose growing strains, the zygotes were subsequently transferred to YP medium containing D-xylose as a sole carbon source. The D-xylose growing cells were inoculated into the liquid fraction of acid pretreated spruce hydrolysate at three different concentrations (40%, 50% and 60%). To maintain the D-xylose growth phenotype, 40 g/L D-xylose was added to the spruce hydrolysate medium. Cells growing in the highest concentration of hydrolysate were re-grown in YP medium containing D-xylose and subsequently used to start the evolutionary adaptation by sequential batch cultivation.

Evolutionary adaptation

The evolutionary adaptation experiment was performed in series of batch fermentations in YP medium containing 40 to 100 g/L D-xylose as a carbon source. Each batch was inoculated with cell pellets obtained from the end of the previous culture, with an initial cell density of OD_{600} of 5. The fermentation was conducted in 100 mL medium in 150 mL cylindrical fermentation tubes at a stirring rate of 120 rpm, and temperature of 35°C. The fermentation tubes were fitted with cotton plugged stoppers to allow CO_2 release.

Determination of mating type by PCR and pheromone assay

The determination of the mating type was done by PCR and pheromone assay. PCR was performed with a primer for the MAT locus and a *MAT**a*** or *MATα* specific primer (Huxley *et al.,* 1990). To validate mating type by a pheromone assay, two tester strains of *S. cerevisiae*, *MAT**a***

bar1-Δ and *MATα sst2-Δ*, were used. A small amount of tester strain was mixed with 1% agar at 50°C and immediately poured on top of a YPD plate. After the top agar solidified, about 10 µL of cell suspension from strains to be tested was spotted onto each tester plate. The cell suspension was prepared by mixing a small amount of cells from a plate in sterile milli-Q water. After 24 h incubation at 30°C, *MATα* cells showed a zone of growth inhibition (halo) on plates of the *bar1-Δ* strain while *MAT**a*** cells showed a zone of growth inhibition on plates of the *sst2-Δ* strain. Diploid cells did not produce a zone of inhibition since they do not produce either of the pheromones.

Molecular Biology methods

Yeast cells were transformed with the LiAc/SS-DNA/PEG method (Gietz et al., 1995) or electroporation modified from Thompson et al., (1998). Genomic DNA from yeast was extracted with PCI [phenol/chlroform/isoamyl-alcohol (25:24:1)] method (Hoffman and Winston, 1987). PCR was performed with Phusion DNA polymerase (New England Biolabs) for construction of the vectors and sequencing purposes, and ExTaq (Takara) for diagnostic purposes. Sanger sequencing was performed by the Genetic Service Facility of the VIB, Belgium.

Plasmid construction

Cloning was performed using the standard restriction and ligation protocol as described in Sambrook *et al.,* (1989). About 1 to 2 µg DNA was digested with the specific restriction enzyme supplied by New England Biolabs, for 2 to 3h at the recommended temperature. Dephosphorylation of the digested vector was performed using FastAP kit from Fermentas life Sciences. Ligation was performed using T4 DNA ligase (Promega) according to the manufacturer's recommendation. *E. coli* cells were transformed using the $CaCl_2$ method (Sambrook *et al.,*1989). Plasmids were propagated in *E. coli* strain TOP10 (Invitrogen), grown in LB medium containing 100 µg/mL ampicillin at 37°C.

Genomic DNA Isolation and pooled segregant whole genome sequencing

All the segregants selected for sequencing were grown in 5mL YPD for 2 days at 30°C. Equal amount of cells from each segregants were pooled based on optical density measurement. The genomic DNA from the pools of segregants and from each parent strain was extracted using the standard protocol described in Johnston (1994). About 6µg high quality DNA samples were sent for sequencing to BGI HONG KONG CO.,LIMITED (Hong Kong). Paired end sequencing was conducted using high-throughput Illumina sequencing technology. A paired end sequence library of 500bp was constructed and sequence reads of 90bp were generated. Average sequence coverage of about 40X was achieved for both strains.

The sequencing reads provided from BGI were aligned onto the reference S288c genome sequence using CLC Genomics Workbench5 or Lasergene's SeqMan Pro software (DNASTAR). The sequencing depth was calculated based on the alignment.

Determination of ploidy by flow cytometry.

Flow cytometric analysis of DNA content was performed according to Popolo *et al.,* (1982). Briefly, exponentially growing cells were washed with ice-cold sterile water and fixed with 70% ethanol. Cells were treated with RNase (1 mg/mL) and the DNA was stained with propidium iodide (0.046 M) in 50 mM Tris, pH 7.7 and 15 mM $MgCl_2$, at 4°C for about 48 h. The fluorescence intensity was measured using a FACScan instrument (Becton Dickinson)

Southern blot analysis

Genomic DNA digested with the appropriate restriction enzyme was run on 0.8% agarose gel overnight at 50 V. Specific probe was prepared by PCR amplification from genomic DNA. The probe was labeled using Amersham Gene Images™ AlkPhos Direct™ labeling and detection system (GE Healthcare). The labeled probe was immediately used to hybridize the DNA that was blotted on a nylon membrane. Chemifluorescent signal was generated and detected using CDP-Star™ as a substrate in conjugation with LAS-4000 luminescent image analyzer.

Reciprocal-hemizygosity analysis (RHA)

RHA was performed according to Steinmetz *et al.,* (2002), in a diploid strain background that carried heterozygous allele for the specific gene of interest. To perform the gene deletions, the KanMX cassette was first amplified from the vector pJET 1,2 B-kanMx-P using primers that contain about 60 bp extra sequences that are homologous to upstream (in the forward primer) and downstream (in the reverse primer) of the gene to be deleted. The PCR product was purified from agarose gel using Wizard® SV Gel and PCR Clean-Up System (Promega) and transformed into the yeast strain. The correct integration of the marker was evaluated by PCR, and the region was subsequently sequenced to confirm the deletion of either of the allele. Two isogenic strains that carry either the mutant or the wild type allele were selected for evaluation.

Determination of specific D-xylose isomerase activity

The specific activity of D-xylose isomerase was measured based on the isomerization of D-xylose to xylulose, followed by reduction of xylulose to xylitol by sorbitol Dehydrogenase (Kersters-Hilderson *et al.,* 1987). Cell extraction was performed by disruption of cells with glass beads in 20mM Tris HCl, pH 8, using a Fast Prep homogenizer. Protein concentration was determined using the Pierce 660nm Protein Assay kit (Thermo Scientific) according to the manufacturer's manual. XI activity in the fresh cell extract was determined at 30°C. The assay mixture contained 100mM Tris-HCl buffer (pH 7.5), 10mM $MgCl_2$, 0.15 mM NADH and 2U sorbitol dehydrogenase. The reaction was started by addition of D-xylose to a final

concentration of 500 mM. A molar extinction coefficient of 6.25 $(mM\ cm)^{-1}$ at 340nm for NADH was used to calculate specific activity. Specific activity was expressed as Units per mg protein. One unit corresponds to the conversion of 1µM of substrate into product in one minute under the specified reaction conditions.

Small-scale fermentations

Semi-anaerobic sequential batch fermentations were performed in 100 mL YP medium containing 40 to 100 g/L D-xylose as sole carbon source, in cylindrical tubes with cotton plugged rubber stopper. Cultures were continuously stirred magnetically at 120 rpm and incubated at 35°C. Anaerobic batch fermentations in synthetic or complete medium were performed in 300 mL shake flasks with a working volume of 200 mL at 35°C. Flasks were closed with fermentation locks containing glycerol. Nitrogen was sparged after cell inoculation until the Oxygen concentration dropped below 2ppm. Cultures were continuously stirred at 120 rpm using a magnetic stirrer. Samples were taken every few hours with needles.

Origin and preparation of lignocellulose hydrolysates

Three different pretreated lignocellulosic biomass materials (*Arundo donax*, spruce tree and a 50/50 mixture of wheat straw and hay), used to evaluate the fermentation performance of the final strain, were obtained from Chemtex (Tortona AL, Italy), SEKAB E-Technology AB (Örnsköldsvik, Sweden), and KaHo Sint-Lieven (Ghent, Belgium), respectively. Pretreatment of *Arundo donax* and spruce were performed by pure steam explosion and SO_2 impregnated steam explosion, respectively. The wheat straw/hay mixture was pretreated using a 0.4 M NaOH at 25°C for 24 hours, after which the residue was washed 3 times with RO water. The *Arundo donax* and spruce pretreated materials were hydrolyzed at an initial pH of 4.8, using enzyme complex ACCELLERASE® 1500 for 48 h at 53°C according to the protocol from the manufacturer. Enzyme hydrolysis of pretreated wheat straw/hay mixture was done using Novozymes cellulose complex (NS50013) and beta-glucosidase (NS50010) at 50°C, pH4.5 for 24 h at KaHo Sint-Lieven (Ghent, Belgium). The fermentations were done at a solid loading of 12% (w/v) for spruce and *Arundo donax* hydrolysate, and 19% (w/v) for wheat straw/hay hydrolysate. For selection of xylose growing and inhibitor tolerant strains after genome shuffling step, pretreated spruce material was used before the enzymatic hydrolysis step.

Metabolite and substrate Analysis

Metabolites and substrate content were analyzed using high performance liquid chromatography (HPLC). Samples from hydrolysates or synthetic medium were centrifuged in 2 mL eppendorf tubes at 14,000 rpm for 5 min. The supernatants were filtered using 0.2 µm filters, and the filtered samples were analyzed immediately or stored at -20 °C.

Metabolites and substrates for fermentation experiment in synthetic or complete medium were analyzed by Waters Isocratic Breeze HPLC system using ion-exchange column WAT010290 and a refractive index detection system (Waters 2414 RI detector, Waters, Milford, MA, USA). Column temperature was maintained at 75°C and 5 mM H_2SO_4 was used as eluent at a flow rate of 1 mL/min.

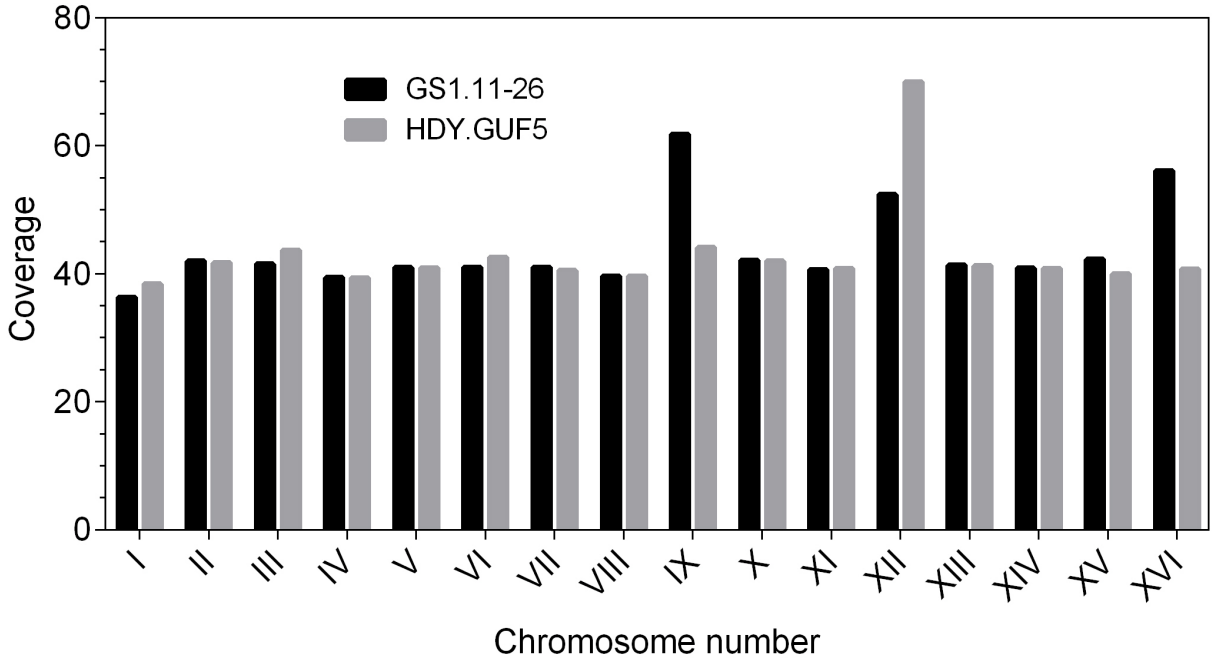
Cell mass concentration

Optical Density (OD_{600nm}) was used to estimate cell dry weight (DW). The DW for inoculums was measured by filtering a 10 mL culture aliquot of a know OD_{600} value in pre-weighed 0.2 mm Supor Membrane disc filters (PALL Corporation, USA), washing the filter with MilliQ water, and drying it in a microwave oven at about 150 watt for 15-20 min to constant weight. The correlation between dry weight and OD_{600} was measured for each strain tested.

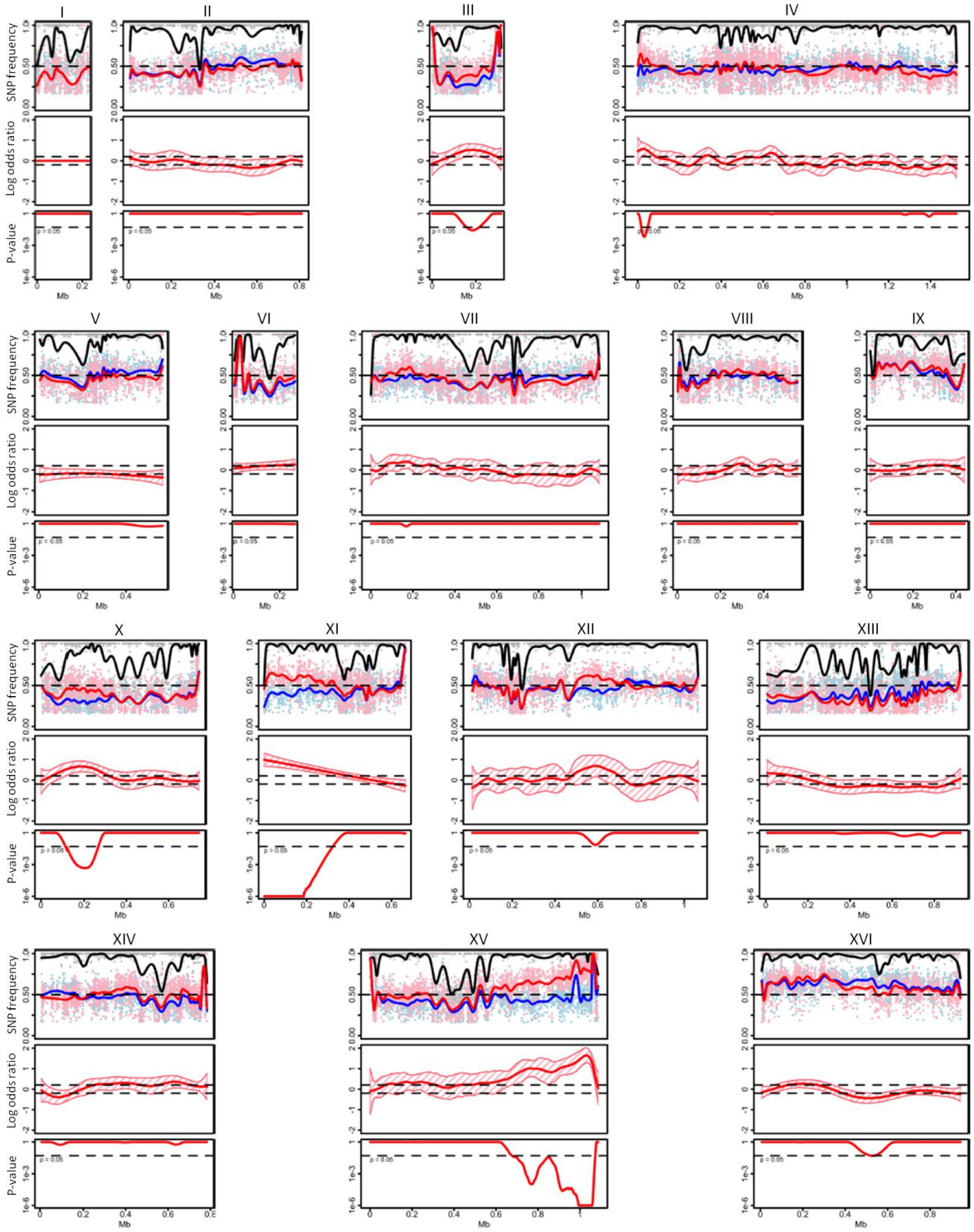
Calculations

Ethanol yield was calculated based on the total amounts of fermentable sugar monomers in the hydrolysates, or the laboratory medium. The specific xylose consumption rate was calculated according to (Zaldivar *et al.,* 2005). For calculation of sugar consumption rate and ethanol productivity in fermentation experiment of in hydrolysates, the initial inoculum biomass was used.

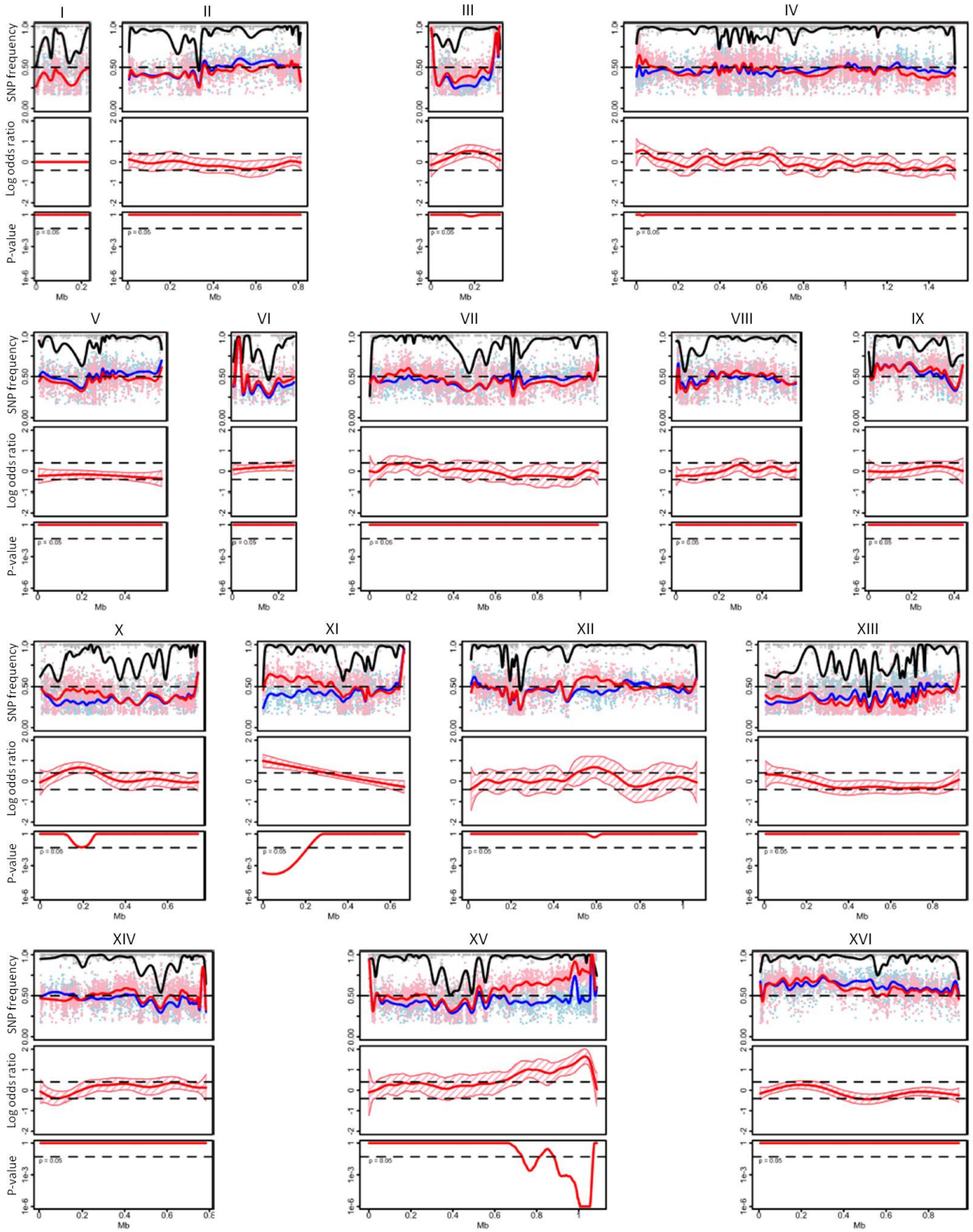
Appendix I. Average genome sequence coverage per chromosome for GS1.11-26 and HDY.GUF5.



Appendix II. Comparison between SNP variant frequency of inferior pool (blue) and superior pool (red) using less stringent analysis, taking log odds ratio outside the range +0.2 and -0.2.



Appendix III. Comparison between SNP variant frequency of inferior pool (blue) and superior pool (red) using more stringent analysis, taking log odds ratio outside the range +0.4 and -0.4.



References

Achten, W.M., Mathijs, E., Verchot, L., Singh, V.P., Aerts, R., Muys, B., 2007. Jatropha biodiesel fueling sustainability? Biofuels, Bioproducts and Biorefining 1, 283–291.

Agbogbo, F., Coward-Kelly, G., 2008. Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast, <i>Pichia stipitis</i> Biotechnology Letters 30, 1515–1524.

Agbor, V.B., Cicek, N., Sparling, R., Berlin, A., Levin, D.B., 2011. Biomass pretreatment: Fundamentals toward application. Biotechnology Advances 29, 675–685.

Agrawal, M., Mao, Z., Chen, R.R., 2011. Adaptation yields a highly efficient xylose-fermenting Zymomonas mobilis strain. Biotechnology and Bioengineering 108, 777–785.

Agrawal, M., Wang, Y., Chen, R.R., 2012. Engineering efficient xylose metabolism into an acetic acid-tolerant Zymomonas mobilis strain by introducing adaptation-induced mutations. Biotechnology letters.

Aguilera, J., Prieto, J., 2001. The Saccharomyces cerevisiae aldose reductase is implied in the metabolism of methylglyoxal in response to stress conditions. Current Genetics 39, 273–283.

Akoh, C.C., Chang, S.-W., Lee, G.-C., Shaw, J.-F., 2007. Enzymatic Approach to Biodiesel Production. J. Agric. Food Chem. 55, 8995–9005.

Albertin, W., Marullo, P., Aigle, M., Bourgais, A., Bely, M., Dillmann, C., De Vienne, D., Sicard, D., 2009. Evidence for autotetraploidy associated with reproductive isolation in Saccharomyces cerevisiae: towards a new domesticated species. Journal of Evolutionary Biology 22, 2157–2170.

Almario, M.P., Reyes, L.H., Kao, K.C., 2013a. Evolutionary engineering of Saccharomyces cerevisiae for enhanced tolerance to hydrolysates of lignocellulosic biomass. Biotechnology and Bioengineering n/a–n/a.

Almario, M.P., Reyes, L.H., Kao, K.C., 2013b. Evolutionary engineering of Saccharomyces cerevisiae for enhanced tolerance to hydrolysates of lignocellulosic biomass. Biotechnology and Bioengineering n/a–n/a.

Almeida, J.R., Modig, T., Petersson, A., Hähn-Hägerdal, B., Lidén, G., Gorwa-Grauslund, M.F., 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. Journal of Chemical Technology & Biotechnology 82, 340–349.

Alper, H., Moxley, J., Nevoigt, E., Fink, G.R., Stephanopoulos, G., 2006. Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production. Science 314, 1565–1568.

Araya, C.L., Payen, C., Dunham, M.J., Fields, S., 2010. Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomics 11, 88.

Ask, M., Olofsson, K., Di Felice, T., Ruohonen, L., Penttilä, M., Lidén, G., Olsson, L., 2012. Challenges in enzymatic hydrolysis and fermentation of pretreated Arundo donax revealed by a comparison between SHF and SSF. Process Biochemistry 47, 1452–1459.

Attfield, P.V., 1997. Stress tolerance: The key to effective strains of industrial baker's yeast. Nat Biotech 15, 1351–1357.

Bai, F.W., Anderson, W.A., Moo-Young, M., 2008. Ethanol fermentation technologies from sugar and starch feedstocks. Biotechnology Advances 26, 89–105.

Bailey, J.E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W.A., Tsai, P.S., 2002. Inverse metabolic engineering: A strategy for directed genetic engineering of useful phenotypes. Biotechnology and Bioengineering 79, 568–579.

Balat, M., Balat, H., 2009. Recent trends in global production and utilization of bio-ethanol fuel. Applied Energy 86, 2273–2282.

Ballesteros, I., Negro, M.J., Oliva, J.M., Cabañas, A., Manzanares, P., Ballesteros, M., 2006. Ethanol production from steam-explosion pretreated wheat straw. Appl. Biochem. Biotechnol. 129-132, 496–508.

Banerjee, S., Archana, A., Satyanarayana, T., 1994. Xylose metabolism in a thermophilic mouldMalbranchea pulchella var.sulfurea TMD-8. Current Microbiology 29, 349–352.

Bellido, C., Bolado, S., Coca, Mã[3]., Lucas, S., GonzÃ¡lez-Benito, G., GarcÃ-a-Cubero, M.T., 2011. Effect of inhibitors formed during wheat straw pretreatment on ethanol fermentation by Pichia stipitis. Bioresource Technology 102, 10868–10874.

Bellissimi, E., Van Dijken, J.P., Pronk, J.T., Van Maris, A.J.A., 2009. Effects of acetic acid on the kinetics of xylose fermentation by an engineered, xylose-isomerase-based Saccharomyces cerevisiae strain. FEMS Yeast Research 9, 358–364.

Bengtsson, O., Hahn-Hagerdal, B., Gorwa-Grauslund, M.F., 2009. Xylose reductase from Pichia stipitis with altered coenzyme preference improves ethanolic xylose fermentation by recombinant Saccharomyces cerevisiae. Biotechnol Biofuels 2, 9.

Bertilsson, M., Olofsson, K., Lidén, G., 2009. Prefermentation improves xylose utilization in simultaneous saccharification and co-fermentation of pretreated spruce. Biotechnology for Biofuels 2, 8.

Bettiga, M., Hahn-Hägerdal, B., Gorwa-Grauslund, M.F., 2008. Comparing the xylose reductase/xylitol dehydrogenase and xylose isomerase pathways in arabinose and xylose fermenting Saccharomyces cerevisiae strains. Biotechnology for Biofuels 1, 16.

Bidlack, J., Malone, M., Benson, R., 1992. Molecular structure and component integration of secondary cell walls in plants. Proceedings of the Oklahoma Academy of Science. 72, 51.

Borneman, A.R., Desany, B.A., Riches, D., Affourtit, J.P., Forgan, A.H., Pretorius, I.S., Egholm, M., Chambers, P.J., 2011. Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of Saccharomyces cerevisiae. PLoS Genet 7, e1001287.

Borneman, A.R., Pretorius, I.S., Chambers, P.J., 2013. Comparative genomics: a revolutionary tool for wine yeast strain development. Current Opinion in Biotechnology 24, 192–199.

Brat, D., Boles, E., Wiedemann, B., 2009. Functional Expression of a Bacterial Xylose Isomerase in Saccharomyces cerevisiae. Applied and Environmental Microbiology 75, 2304–2311.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., Ahn, J., Dewar-Darch, D., Reguly, T., Tang, X., Almeida, R., Qin, Z.S., Pawson, T., Gingras, A.-C., Nesvizhskii, A.I., Tyers, M., 2010. A Global Protein Kinase and Phosphatase Interaction Network in Yeast. Science 328, 1043–1046.

Bruinenberg, P.M., Bot, P.H.M., Dijken, J.P., Scheffers, W.A., 1983. The role of redox balances in the anaerobic fermentation of xylose by yeasts. European Journal of Applied Microbiology and Biotechnology 18, 287–292.

Cagney, G., Uetz, P., Fields, S., 2001. Two-hybrid analysis of the Saccharomyces cerevisiae 26S proteasome. Physiol. Genomics 7, 27–34.

Cai, Z., Zhang, B., Li, Y., 2012. Engineering Saccharomyces cerevisiae for efficient anaerobic xylose fermentation: Reflections and perspectives. Biotechnology Journal 7, 34–46.

Chandra, R.P., Bura, R., Mabee, W.E., Berlin, A., Pan, X., Saddler, J.N., 2007. Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? Adv. Biochem. Eng. Biotechnol. 108, 67–93.

Chang, V.S., Burr, B., Holtzapple, M.T., 1997. Lime pretreatment of switchgrass. Appl. Biochem. Biotechnol. 63-65, 3–19.

Charles, J.S., Hamilton, M.L., Petes, T.D., 2010. Meiotic Chromosome Segregation in Triploid Strains of Saccharomyces cerevisiae. Genetics 186, 537–550.

Chisti, Y., 2008. Biodiesel from microalgae beats bioethanol. Trends in Biotechnology 26, 126–131.

Chu, B.C.H., Lee, H., 2007. Genetic improvement of Saccharomyces cerevisiae for xylose fermentation. Biotechnology Advances 25, 425–441.

Claesen, J., Clement, L., Shkedy, Z., Foulquié-Moreno, M.R., Burzykowski, T., 2013. Simultaneous Mapping of Multiple Gene Loci with Pooled Segregants. PLoS ONE 8, e55133.

Ding, M.-Z., Li, B.-Z., Cheng, J.-S., Yuan, Y.-J., 2010. Metabolome analysis of differential responses of diploid and haploid yeast to ethanol stress. OMICS 14, 553–561.

Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., Botstein, D., 2002. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. PNAS 99, 16144–16149.

Elena, S.F., Lenski, R.E., 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4, 457–469.

Eliasson, A., Christensson, C., Wahlbom, C.F., Hahn-Hagerdal, B., 2000. Anaerobic Xylose Fermentation by Recombinant Saccharomyces cerevisiae Carrying XYL1, XYL2, and XKS1 in Mineral Medium Chemostat Cultures. Appl. Environ. Microbiol. 66, 3381–3386.

Fairley, P., 2011. Introduction: Next generation biofuels. Nature 474, S2–S5.

Fiaux, J., Çakar, Z.P., Sonderegger, M., Wüthrich, K., Szyperski, T., Sauer, U., 2003. Metabolic-Flux Profiling of the Yeasts Saccharomyces cerevisiae and Pichia stipitis. Eukaryot Cell 2, 170–180.

Galbe, M., Zacchi, G., 2007. Pretreatment of lignocellulosic materials for efficient bioethanol production. Adv. Biochem. Eng. Biotechnol. 108, 41–65.

Galeote, V., Bigey, F., Beyne, E., Novo, M., Legras, J.-L., Casaregola, S., Dequin, S., 2011. Amplification of a Zygosaccharomyces bailii DNA Segment in Wine Yeast Genomes by Extrachromosomal Circular DNA Formation. PLoS ONE 6, e17872.

Gárdonyi, M., Hahn-Hägerdal, B., 2003. The Streptomyces rubiginosus xylose isomerase is misfolded when expressed in Saccharomyces cerevisiae. Enzyme and Microbial Technology 32, 252–259.

Gerstein, A.C., Chun, H.-J.E., Grant, A., Otto, S.P., 2006. Genomic Convergence toward Diploidy in Saccharomyces cerevisiae. PLoS Genet 2, e145.

Gietz, R.D., Schiestl, R.H., Willems, A.R., Woods, R.A., 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast 11, 355–360.

Goldemberg, J., 2007. Ethanol for a Sustainable Energy Future. Science 315, 808–810.

González-Fernández, C., Molinuevo-Salces, B., García-González, M.C., 2011. Nitrogen transformations under different conditions in open ponds by means of microalgae–bacteria consortium treating pig slurry. Bioresource Technology 102, 960–966.

Gorsich, S.W., Dien, B.S., Nichols, N.N., Slininger, P.J., Liu, Z.L., Skory, C.D., 2006. Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in Saccharomyces cerevisiae. Appl Microbiol Biotechnol 71, 339–349.

Gresham, D., Usaite, R., Germann, S.M., Lisby, M., Botstein, D., Regenberg, B., 2010. Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. PNAS 107, 18551–18556.

Hahn-Hägerdal, B., Karhumaa, K., Fonseca, C., Spencer-Martins, I., Gorwa-Grauslund, M.F., 2007a. Towards industrial pentose-fermenting yeast strains. Appl. Microbiol. Biotechnol. 74, 937–953.

Hahn-Hägerdal, B., Karhumaa, K., Jeppsson, M., Gorwa-Grauslund, M.F., 2007b. Metabolic engineering for pentose utilization in Saccharomyces cerevisiae. Adv. Biochem. Eng. Biotechnol. 108, 147–177.

Hamacher, T., Becker, J., Gárdonyi, M., Hahn-Hägerdal, B., Boles, E., 2002. Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. Microbiology 148, 2783–2788.

Harhangi, H.R., Akhmanova, A.S., Emmens, R., van der Drift, C., de Laat, W.T.A.M., van Dijken, J.P., Jetten, M.S.M., Pronk, J.T., Op den Camp, H.J.M., 2003. Xylose metabolism in the anaerobic fungus <i>Piromyces</i> sp. strain E2 follows the bacterial pathway. Archives of Microbiology 180, 134–141.

Hashimoto, S., Aritomi, K., Minohara, T., Nishizawa, Y., Hoshida, H., Kashiwagi, S., Akada, R., 2006. Direct mating between diploid sake strains of Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. 69, 689–696.

Hector, R., Qureshi, N., Hughes, S., Cotta, M., 2008. Expression of a heterologous xylose transporter in a <i>Saccharomyces cerevisiae</i> strain engineered to utilize xylose improves aerobic xylose consumption. Applied Microbiology and Biotechnology 80, 675–684.

Hector, R.E., Dien, B.S., Cotta, M.A., Qureshi, N., 2010. Engineering industrial Saccharomyces cerevisiae strains for xylose fermentation and comparison for switchgrass conversion. Journal of Industrial Microbiology & Biotechnology 38, 1193–1202.

Ho, N.W., Chen, Z., Brainard, A.P., 1998. Genetically engineered Saccharomyces yeast capable of effective cofermentation of glucose and xylose. Appl. Environ. Microbiol. 64, 1852–1859.

Hong, K.-K., Nielsen, J., 2012. Recovery of phenotypes obtained by adaptive evolution through inverse metabolic engineering. Appl. Environ. Microbiol.

Hoogwijk, M., Faaij, A., van den Broek, R., Berndes, G., Gielen, D., Turkenburg, W., 2003. Exploration of the ranges of the global potential of biomass for energy. Biomass and Bioenergy 25, 119–133.

Hou, L., 2009. Novel methods of genome shuffling in Saccharomyces cerevisiae. Biotechnology Letters 31, 671–677.

Hou, X., Yao, S., 2012. Improved inhibitor tolerance in xylose-fermenting yeast Spathaspora passalidarum by mutagenesis and protoplast fusion. Appl. Microbiol. Biotechnol. 93, 2591–2601.

Huxley, C., Green, E.D., Dunham, I., 1990. Rapid assessment of S. cerevisiae mating type by PCR. Trends Genet. 6, 236.

Ingram, L.O., Conway, T., Clark, D.P., Sewell, G.W., Preston, J.F., 1987. Genetic engineering of ethanol production in Escherichia coli. Appl Environ Microbiol 53, 2420–2425.

Jeffries, T.W., 2006. Engineering yeasts for xylose metabolism. Current Opinion in Biotechnology 17, 320–326.

Jeffries, T.W., Jin, Y.S., 2000. Ethanol and thermotolerance in the bioconversion of xylose by yeasts. Adv. Appl. Microbiol. 47, 221–268.

Jeppsson, M., Johansson, B., Jensen, P.R., Hahn-Hägerdal, B., Gorwa-Grauslund, M.F., 2003. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant Saccharomyces cerevisiae strains. Yeast 20, 1263–1272.

Jin, Y.-S., Jeffries, T.W., 2003. Changing flux of xylose metabolites by altering expression of xylose reductase and xylitol dehydrogenase in recombinant Saccharomyces cerevisiae. Appl. Biochem. Biotechnol. 105 -108, 277–286.

Jin, Y.-S., Ni, H., Laplaza, J.M., Jeffries, T.W., 2003. Optimal Growth and Ethanol Production from Xylose by Recombinant Saccharomyces cerevisiae Require Moderate d-Xylulokinase Activity. Appl Environ Microbiol 69, 495–503.

Jingping, G., Hongbing, S., Gang, S., Hongzhi, L., Wenxiang, P., 2012. A genome shuffling-generated Saccharomyces cerevisiae isolate that ferments xylose and glucose to produce high levels of ethanol. Journal of Industrial Microbiology & Biotechnology.

Johansson, B., Christensson, C., Hobley, T., Hahn-Hägerdal, B., 2001. Xylulokinase Overexpression in Two Strains of Saccharomyces cerevisiae Also Expressing Xylose Reductase and Xylitol Dehydrogenase and Its Effect on Fermentation of Xylose and Lignocellulosic Hydrolysate. Appl Environ Microbiol 67, 4249–4255.

Jönsson, L.J., Alriksson, B., Nilvebrant, N.-O., 2013. Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnology for Biofuels 6, 16.

Johnston JR., 1994. Molecular genetics of yeast: A practical approach. Oxford University Press, New York.

Jørgensen, H., Kristensen, J.B., Felby, C., 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioproducts and Biorefining 1, 119–134.

Juan, J.C., Kartika, D.A., Wu, T.Y., Hin, T.-Y.Y., 2011. Biodiesel production from jatropha oil by catalytic and non-catalytic approaches: An overview. Bioresource Technology 102, 452–460.

Jun, H., Kieselbach, T., Jönsson, L.J., 2011. Enzyme production by filamentous fungi: analysis of the secretome of Trichoderma reesei grown on unconventional carbon source. Microb Cell Fact 10, 68.

Karhumaa, K., Fromanger, R., Hahn-Hägerdal, B., Gorwa-Grauslund, M.-F., 2007a. High activity of xylose reductase and xylitol dehydrogenase improves xylose fermentation by recombinant <i>Saccharomyces cerevisiae</i> Applied Microbiology and Biotechnology 73, 1039–1046.

Karhumaa, K., Hahn-Hägerdal, B., Gorwa-Grauslund, M.-F., 2005. Investigation of limiting metabolic steps in the utilization of xylose by recombinant Saccharomyces cerevisiae using metabolic engineering. Yeast 22, 359–368.

Karhumaa, K., Sanchez, R., Hahn-Hagerdal, B., Gorwa-Grauslund, M.-F., 2007b. Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant Saccharomyces cerevisiae. Microbial Cell Factories 6, 5.

Karlsson, J., Saloheimo, M., Siika-Aho, M., Tenkanen, M., Penttilä, M., Tjerneld, F., 2001. Homologous expression and characterization of Cel61A (EG IV) of Trichoderma reesei. Eur. J. Biochem. 268, 6498–6507.

Kersters-Hilderson, H., Callens, M., Van Opstal, O., Vangrysperre, W., De Bruyne, C.K., 1987. Kinetic characterization of d-xylose isomerases by enzymatic assays using d-sorbitol dehydrogenase. Enzyme and Microbial Technology 9, 145–148.

Kim, S., Dale, B.E., 2004. Global potential bioethanol production from wasted crops and crop residues. Biomass and Bioenergy 26, 361–375.

Koppram, R., Albers, E., Olsson, L., 2012. Evolutionary engineering strategies to enhance tolerance of xylose utilizing recombinant yeast to inhibitors derived from spruce biomass. Biotechnology for Biofuels 5, 32.

Kuhn, A., Zyl, C. van, Tonder, A. van, Prior, B.A., 1995. Purification and partial characterization of an aldo-keto reductase from Saccharomyces cerevisiae. Appl. Environ. Microbiol. 61, 1580–1585.

Kuyper, M., Harhangi, H.R., Stave, A.K., Winkler, A.A., Jetten, M.S.M., de Laat, W.T.A.M., den Ridder, J.J.J., Op den Camp, H.J.M., van Dijken, J.P., Pronk, J.T., 2003a. High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by Saccharomyces cerevisiae? FEMS Yeast Research 4, 69–78.

Kuyper, M., Harhangi, H.R., Stave, A.K., Winkler, A.A., Jetten, M.S.M., de Laat, W.T.A.M., den Ridder, J.J.J., Op den Camp, H.J.M., van Dijken, J.P., Pronk, J.T., 2003b. High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by Saccharomyces cerevisiae? FEMS Yeast Research 4, 69–78.

Kuyper, M., Hartog, M.M.P., Toirkens, M.J., Almering, M.J.H., Winkler, A.A., van Dijken, J.P., Pronk, J.T., 2005a. Metabolic engineering of a xylose-isomerase-expressing Saccharomyces cerevisiae strain for rapid anaerobic xylose fermentation. FEMS Yeast Research 5, 399–409.

Kuyper, M., Toirkens, M.J., Diderich, J.A., Winkler, A.A., van Dijken, J.P., Pronk, J.T., 2005b. Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting Saccharomyces cerevisiae strain. FEMS Yeast Research 5, 925–934.

Kuyper, M., Winkler, A.A., van Dijken, J.P., Pronk, J.T., 2004. Minimal metabolic engineering of Saccharomyces cerevisiae for efficient anaerobic xylose fermentation: a proof of principle. FEMS Yeast Research 4, 655–664.

Larsson, Simona, Cassland, P., Jonsson, L.J., 2001. Development of a Saccharomyces cerevisiae Strain with Enhanced Resistance to Phenolic Fermentation Inhibitors in Lignocellulose Hydrolysates by Heterologous Expression of Laccase. Appl Environ Microbiol 67, 1163–1170.

Larsson, S, Nilvebrant, N.O., Jönsson, L.J., 2001. Effect of overexpression of Saccharomyces cerevisiae Pad1p on the resistance to phenylacrylic acids and lignocellulose hydrolysates under aerobic and oxygen-limited conditions. Appl. Microbiol. Biotechnol. 57, 167–174.

Lau, M.W., Gunawan, C., Balan, V., Dale, B.E., 2010. Comparing the fermentation performance of Escherichia coli KO11, Saccharomyces cerevisiae 424A(LNH-ST) and Zymomonas mobilis AX101 for cellulosic ethanol production. Biotechnology for Biofuels 3, 11.

Leandro, M.J., Gonçalves, P., Spencer-Martins, I., 2006. Two glucose/xylose transporter genes from the yeast Candida intermedia: first molecular characterization of a yeast xylose–H+ symporter. Biochem J 395, 543–549.

Lehr, F., Posten, C., 2009. Closed photo-bioreactors as tools for biofuel production. Current Opinion in Biotechnology 20, 280–285.

Liang, L., Zhang, J., Lin, Z., 2007. Altering coenzyme specificity of Pichia stipitis xylose reductase by the semi-rational approach CASTing. Microbial Cell Factories 6, 36.

Liu, E., Hu, Y., 2010. Construction of a xylose-fermenting Saccharomyces cerevisiae strain by combined approaches of genetic engineering, chemical mutagenesis and evolutionary adaptation. Biochemical Engineering Journal 48, 204–210.

Liu, J.-J., Ding, W.-T., Zhang, G.-C., Wang, J.-Y., 2011. Improving ethanol fermentation performance of Saccharomyces cerevisiae in very high-gravity fermentation through chemical mutagenesis and meiotic recombination. Applied Microbiology and Biotechnology 91, 1239–1246.

Lönn, A., Gárdonyi, M., van Zyl, W., Hahn-Hägerdal, B., Otero, R.C., 2002. Cold adaptation of xylose isomerase from Thermus thermophilus through random PCR mutagenesis. Gene cloning and protein characterization. Eur. J. Biochem. 269, 157–163.

Mackie, K.L., Brownell, H.H., West, K.L., Saddler, J.N., 1985. Effect of Sulphur Dioxide and Sulphuric Acid on Steam Explosion of Aspenwood. Journal of Wood Chemistry and Technology 5, 405–425.

Macrelli, S., Mogensen, J., Zacchi, G., 2012. Techno-economic evaluation of 2nd generation bioethanol production from sugar cane bagasse and leaves integrated with the sugar-based ethanol process. Biotechnology for Biofuels 5, 22.

Madhavan, A., Tamalampudi, S., Srivastava, A., Fukuda, H., Bisaria, V.S., Kondo, A., 2009. Alcoholic fermentation of xylose and mixed sugars using recombinant Saccharomyces cerevisiae engineered for xylose utilization. Appl Microbiol Biotechnol 82, 1037–1047.

Marullo, P., Mansour, C., Dufour, M., Albertin, W., Sicard, D., Bely, M., Dubourdieu, D., 2009. Genetic improvement of thermo-tolerance in wine Saccharomyces cerevisiae strains by a backcross approach. FEMS Yeast Research 9, 1148–1160.

Matsushika, A., Goshima, T., Fujii, T., Inoue, H., Sawayama, S., Yano, S., 2012. Characterization of non-oxidative transaldolase and transketolase enzymes in the pentose phosphate pathway with regard to xylose utilization by recombinant Saccharomyces cerevisiae. Enzyme and Microbial Technology 51, 16–25.

Matsushika, A., Inoue, H., Watanabe, S., Kodaki, T., Makino, K., Sawayama, S., 2009. Efficient Bioethanol Production by a Recombinant Flocculent Saccharomyces cerevisiae Strain with a Genome-

Integrated NADP+-Dependent Xylitol Dehydrogenase Gene. Appl Environ Microbiol 75, 3818–3822.

Menetrez, M.Y., 2012. An Overview of Algae Biofuel Production and Potential Environmental Impact. Environ. Sci. Technol. 46, 7073–7085.

Michel, A.H., Kornmann, B., Dubrana, K., Shore, D., 2005. Spontaneous rDNA copy number variation modulates Sir2 levels and epigenetic gene silencing. Genes Dev 19, 1199–1210.

Mielenz, J.R., 2001. Ethanol production from biomass: technology and commercialization status. Current Opinion in Microbiology 4, 324–329.

Modi, M.K., Reddy, J.R.C., Rao, B.V.S.K., Prasad, R.B.N., 2007. Lipase-mediated conversion of vegetable oils into biodiesel using ethyl acetate as acyl acceptor. Bioresource Technology 98, 1260–1264.

Mohammad J. Taherzadeh, K.K., 2007. Acid-based hydrolysis processes for ethanol from lignocellulosic materials: A review. BioResources.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y.Y., Holtzapple, M., Ladisch, M., 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresource Technology 96, 673–686.

Mutturi, S., Lidén, G., 2013. Effect of Temperature on Simultaneous Saccharification and Fermentation of Pretreated Spruce and Arundo. Ind. Eng. Chem. Res. 52, 1244–1251.

Nevoigt, E., 2008. Progress in Metabolic Engineering of Saccharomyces cerevisiae. Microbiol Mol Biol Rev 72, 379–412.

Nguyen, Q.A., Tucker, M.P., Keller, F.A., Eddy, F.P., 2000. Two-stage dilute-acid pretreatment of softwoods. Appl. Biochem. Biotechnol. 84-86, 561–576.

Nieduszynski, C.A., Knox, Y., Donaldson, A.D., 2006. Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev. 20, 1874–1879.

Nissen, T.L., Anderlund, M., Nielsen, J., Villadsen, J., Kielland-Brandt, M.C., 2001. Expression of a cytoplasmic transhydrogenase in Saccharomyces cerevisiae results in formation of 2-oxoglutarate due to depletion of the NADPH pool. Yeast 18, 19–32.

Olofsson, K., Bertilsson, M., Lidén, G., 2008a. A short review on SSF – an interesting process option for ethanol production from lignocellulosic feedstocks. Biotechnol Biofuels 1, 7.

Olofsson, K., Palmqvist, B., Lidén, G., 2010. Improving simultaneous saccharification and co-fermentation of pretreated wheat straw using both enzyme and substrate feeding. Biotechnology for Biofuels 3, 17.

Olofsson, K., Rudolf, A., Lidén, G., 2008b. Designing simultaneous saccharification and fermentation for improved xylose conversion by a recombinant strain of Saccharomyces cerevisiae. Journal of Biotechnology 134, 112–120.

Olofsson, K., Runquist, D., Hahn-Hägerdal, B., Lidén, G., 2011. A mutated xylose reductase increases bioethanol production more than a glucose/xylose facilitator in simultaneous fermentation and co-fermentation of wheat straw. AMB Express 1, 4.

Pais, T.M., Foulquié-Moreno, M.R., Hubmann, G., Duitama, J., Swinnen, S., Goovaerts, A., Yang, Y., Dumortier, F., Thevelein, J.M., 2013. Comparative Polygenic Analysis of Maximal Ethanol Accumulation Capacity and Tolerance to High Ethanol Levels of Cell Proliferation in Yeast. PLoS Genet 9, e1003548.

Palmqvist, E., Hahn-Hägerdal, B., 2000. Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. Bioresource Technology 74, 25–33.

Parts, L., Cubillos, F.A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S.J., Molin, M., Zia, A., Simpson, J.T., Quail, M.A., Moses, A., Louis, E.J., Durbin, R., Liti, G., 2011. Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res. 21, 1131–1138.

Patnaik, R., 2008. Engineering Complex Phenotypes in Industrial Strains. Biotechnology Progress 24, 38–47.

Pavelka, N., Rancati, G., Zhu, J., Bradford, W.D., Saraf, A., Florens, L., Sanderson, B.W., Hattem, G.L., Li, R., 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature 468, 321–325.

Peng, B., Shen, Y., Li, X., Chen, X., Hou, J., Bao, X., 2012. Improvement of xylose fermentation in respiratory-deficient xylose-fermenting Saccharomyces cerevisiae. Metabolic Engineering 14, 9–18.

Persson, P., Larsson, S., Jönsson, L.J., Nilvebrant, N.-O., Sivik, B., Munteanu, F., Thörneby, L., Gorton, L., 2002. Supercritical fluid extraction of a lignocellulosic hydrolysate of spruce for detoxification and to facilitate analysis of inhibitors. Biotechnology and Bioengineering 79, 694–700.

Petes, T.D., 1980. Molecular Genetics of Yeast. Annual Review of Biochemistry 49, 845–876.

Petri, R., Schmidt-Dannert, C., 2004. Dealing with complexity: evolutionary engineering and genome shuffling. Current Opinion in Biotechnology 15, 298–304.

Phillips, C.M., Beeson, W.T., Cate, J.H., Marletta, M.A., 2011. Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by Neurospora crassa. ACS Chem. Biol. 6, 1399–1406.

Popolo, L., Vanoni, M., Alberghina, L., 1982. Control of the yeast cell cycle by protein synthesis. Experimental Cell Research 142, 69–78.

Prasetyo, J., Naruse, K., Kato, T., Boonchird, C., Harashima, S., Park, E., 2011. Bioconversion of paper sludge to biofuel by simultaneous saccharification and fermentation using a cellulase of paper sludge origin and thermotolerant Saccharomyces cerevisiae TJ14. Biotechnology for Biofuels 4, 35.

Pretorius, I.S., Bauer, F.F., 2002. Meeting the consumer challenge through genetically customized wine-yeast strains. Trends in Biotechnology 20, 426–432.

Randelli, F., 2009. An integrated analysis of production costs and net energy balance of biofuels. Regional Environmental Change 9, 221–229.

Reyes, L.H., Winkler, J., Kao, K.C., 2012. Visualizing evolution in real-time method for strain engineering. Front. Microbio. 3, 198.

Roca, C., Nielsen, J., Olsson, L., 2003. Metabolic Engineering of Ammonium Assimilation in Xylose-Fermenting Saccharomyces cerevisiae Improves Ethanol Production. Appl. Environ. Microbiol. 69, 4732–4736.

Runquist, D., Hahn-Hägerdal, B., Bettiga, M., 2010. Increased Ethanol Productivity in Xylose-Utilizing Saccharomyces cerevisiae via a Randomly Mutagenized Xylose Reductase. Appl Environ Microbiol 76, 7796–7802.

Saloheimo, A., Rauta, J., Stasyk, O., Sibirny, A., Penttilä, M., Ruohonen, L., 2007. Xylose transport studies with xylose-utilizing <i>Saccharomyces cerevisiae</i> strains expressing heterologous and homologous permeases. Applied Microbiology and Biotechnology 74, 1041–1052.

Sambrook J, Fritsch EF, Maniatis T, 1989. Molecular cloning: a laboratory manual, 2nd. ed. Cold Spring Harbor, New York.

Sánchez, Ó.J., Cardona, C.A., 2008. Trends in biotechnological production of fuel ethanol from different feedstocks. Bioresource Technology 99, 5270–5295.

Sanderson, K., 2011. Lignocellulose: A chewy problem. Nature 474, S12–S14.

Sauer, U., 2001. Evolutionary engineering of industrially important microbial phenotypes. Adv. Biochem. Eng. Biotechnol. 73, 129–169.

Schmidt-Dannert, C., Arnold, F.H., 1999. Directed evolution of industrial enzymes. Trends in Biotechnology 17, 135–136.

Schubert, C., 2006. Can biofuels finally take center stage? Nature Biotechnology 24, 777–784.

Schwartz, K., Wenger, J.W., Dunn, B., Sherlock, G., 2012. APJ1 and GRE3 Homologs Work in Concert to Allow Growth in Xylose in a Natural Saccharomyces Sensu Stricto Hybrid Yeast. Genetics.

Schwietzke, S., Kim, Y., Ximenes, E., Mosier, N., Ladisch, M., 2009. Ethanol Production from Maize, in: Kriz, P.D.A.L., Larkins, P.D.B.A. (Eds.), Molecular Genetic Approaches to Maize Improvement, Biotechnology in Agriculture and Forestry. Springer Berlin Heidelberg, pp. 347–364.

Scott, S.A., Davey, M.P., Dennis, J.S., Horst, I., Howe, C.J., Lea-Smith, D.J., Smith, A.G., 2010. Biodiesel from algae: challenges and prospects. Current Opinion in Biotechnology 21, 277–286.

Sedlak, M., Ho, N., 2004a. Production of ethanol from cellulosic biomass hydrolysates using genetically engineered <i>saccharomyces</i> yeast capable of cofermenting glucose and xylose. Applied Biochemistry and Biotechnology 114, 403–416.

Sedlak, M., Ho, N.W.Y., 2004b. Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant Saccharomyces yeast. Yeast 21, 671–684.

Shaw, A.J., Podkaminer, K.K., Desai, S.G., Bardsley, J.S., Rogers, S.R., Thorne, P.G., Hogsett, D.A., Lynd, L.R., 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. PNAS 105, 13769–13774.

Shen, Y., Chen, X., Peng, B., Chen, L., Hou, J., Bao, X., 2012. An efficient xylose-fermenting recombinant Saccharomyces cerevisiae strain obtained through adaptive evolution and its global transcription profile. Applied Microbiology and Biotechnology 96, 1079–1091.

Shimada, Y., Watanabe, Y., Sugihara, A., Tominaga, Y., 2002. Enzymatic alcoholysis for biodiesel fuel production and application of the reaction to oil processing. Journal of Molecular Catalysis B: Enzymatic 17, 133–142.

Shuai, L., Yang, Q., Zhu, J.Y., Lu, F.C., Weimer, P.J., Ralph, J., Pan, X.J., 2010. Comparative study of SPORL and dilute-acid pretreatments of spruce for cellulosic ethanol production. Bioresource Technology 101, 3106–3114.

Sims, R.E.H., Mabee, W., Saddler, J.N., Taylor, M., 2010. An overview of second generation biofuel technologies. Bioresource Technology 101, 1570–1580.

Singh, M.V., Anthony Weil, P., 2002. A method for plasmid purification directly from yeast. Analytical Biochemistry 307, 13–17.

Skoog, K., Hahn-Hägerdal, B., 1990. Effect of Oxygenation on Xylose Fermentation by Pichia stipitis. Appl Environ Microbiol 56, 3389–3394.

Smith, D.C., Forsberg, C.W., 1991. α-Glucuronidase and Other Hemicellulase Activities of Fibrobacter succinogenes S85 Grown on Crystalline Cellulose or Ball-Milled Barley Straw. Appl Environ Microbiol 57, 3552–3557.

Sonderegger, M., Sauer, U., 2003. Evolutionary Engineering of Saccharomyces cerevisiae for Anaerobic Growth on Xylose. Appl Environ Microbiol 69, 1990–1998.

Stift, M., Reeve, R., Van TIENDEREN, P.H., 2010. Inheritance in tetraploid yeast revisited: segregation patterns and statistical power under different inheritance models. Journal of Evolutionary Biology 23, 1570–1578.

Su, E., Wei, D., 2008. Improvement in lipase-catalyzed methanolysis of triacylglycerols for biodiesel production using a solvent engineering method. Journal of Molecular Catalysis B: Enzymatic 55, 118–125.

Subtil, T., Boles, E., 2012. Competition between pentoses and glucose during uptake and catabolism in recombinant Saccharomyces cerevisiae. Biotechnology for Biofuels 5, 14.

Swinnen, S., Schaerlaekens, K., Pais, T., Claesen, J., Hubmann, G., Yang, Y., Demeke, M., Foulquié-Moreno, M.R., Goovaerts, A., Souvereyns, K., Clement, L., Dumortier, F., Thevelein, J.M., 2012a.

Identification of Novel Causative Genes Determining the Complex Trait of High Ethanol Tolerance in Yeast Using Pooled-Segregant Whole-Genome Sequence Analysis. Genome Res.

Swinnen, S., Thevelein, J.M., Nevoigt, E., 2012b. Genetic mapping of quantitative phenotypic traits in Saccharomyces cerevisiae. FEMS Yeast Research 12, 215–227.

Taherzadeh, M.J., Karimi, K., 2008. Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review. Int J Mol Sci 9, 1621–1651.

Tengborg, C., Galbe, M., Zacchi, G., 2001. Reduced inhibition of enzymatic hydrolysis of steam-pretreated softwood. Enzyme and Microbial Technology 28, 835–844.

Teunissen, A., Dumortier, F., Gorwa, M.-F., Bauer, J., Tanghe, A., Loiez, A., Smet, P., Van Dijck, P., Thevelein, J.M., 2002. Isolation and Characterization of a Freeze-Tolerant Diploid Derivative of an Industrial Baker's Yeast Strain and Its Use in Frozen Doughs. Applied and Environmental Microbiology 68, 4780–4787.

Thompson, J.R., Register, E., Curotto, J., Kurtz, M., Kelly, R., 1998. An improved protocol for the preparation of yeast cells for transformation by electroporation. Yeast 14, 565–571.

Toivari, M.H., Aristidou, A., Ruohonen, L., Penttilä, M., 2001. Conversion of xylose to ethanol by recombinant Saccharomyces cerevisiae: importance of xylulokinase (XKS1) and oxygen availability. Metab. Eng. 3, 236–249.

Toivari, M.H., Salusjärvi, L., Ruohonen, L., Penttilä, M., 2004. Endogenous Xylose Pathway in Saccharomyces cerevisiae. Appl Environ Microbiol 70, 3681–3686.

Traff, K.L., Otero Cordero, R.R., van Zyl, W.H., Hahn-Hagerdal, B., 2001. Deletion of the GRE3 Aldose Reductase Gene and Its Influence on Xylose Metabolism in Recombinant Strains of Saccharomyces cerevisiae Expressing the xylA and XKS1 Genes. Appl Environ Microbiol 67, 5668–5674.

Unrean, P., Nguyen, N.H.A., 2012. Metabolic pathway analysis of Scheffersomyces (Pichia) stipitis: effect of oxygen availability on ethanol synthesis and flux distributions. Applied Microbiology and Biotechnology 94, 1387–1398.

Van Maris, A.J.A., Winkler, A.A., Kuyper, M., de Laat, W.T.A.M., van Dijken, J.P., Pronk, J.T., 2007. Development of efficient xylose fermentation in Saccharomyces cerevisiae: xylose isomerase as a key component. Adv. Biochem. Eng. Biotechnol. 108, 179–204.

Van Steen, E., Claeys, M., 2008. Fischer-Tropsch Catalysts for the Biomass-to-Liquid (BTL)-Process. Chemical Engineering & Technology 31, 655–666.

Van Zyl, W.H., Lynd, L.R., den Haan, R., McBride, J.E., 2007. Consolidated bioprocessing for bioethanol production using Saccharomyces cerevisiae. Adv. Biochem. Eng. Biotechnol. 108, 205–235.

Wahlbom, C. Fredrik, Cordero Otero, R.R., van Zyl, W.H., Hahn-Hagerdal, B., Jonsson, L.J., 2003. Molecular Analysis of a Saccharomyces cerevisiae Mutant with Improved Ability To Utilize Xylose Shows Enhanced Expression of Proteins Involved in Transport, Initial Xylose Metabolism, and the Pentose Phosphate Pathway. Appl. Environ. Microbiol. 69, 740–746.

Wahlbom, C.Fredrik, van Zyl, W.H., Jönsson, L.J., Hahn-Hägerdal, B., Otero, R.R.C., 2003. Generation of the improved recombinant xylose-utilizing Saccharomyces cerevisiae TMB 3400 by random mutagenesis and physiological comparison with Pichia stipitis CBS 6054. FEMS Yeast Research 3, 319–326.

Walfridsson, M., Bao, X., Anderlund, M., Lilius, G., Bülow, L., Hahn-Hägerdal, B., 1996. Ethanolic fermentation of xylose with Saccharomyces cerevisiae harboring the Thermus thermophilus xylA gene, which expresses an active xylose (glucose) isomerase. Appl Environ Microbiol 62, 4648–4651.

Watanabe, S., Abu Saleh, A., Pack, S.P., Annaluru, N., Kodaki, T., Makino, K., 2007. Ethanol production from xylose by recombinant Saccharomyces cerevisiae expressing protein-engineered NADH-preferring xylose reductase from Pichia stipitis. Microbiology (Reading, Engl.) 153, 3044–3054.

Watanabe, S., Kodaki, T., Makino, K., 2005. Complete Reversal of Coenzyme Specificity of Xylitol Dehydrogenase and Increase of Thermostability by the Introduction of Structural Zinc. J. Biol. Chem. 280, 10340–10349.

Weber, C., Farwick, A., Benisch, F., Brat, D., Dietz, H., Subtil, T., Boles, E., 2010. Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. Applied Microbiology and Biotechnology 87, 1303–1315.

Wenger, J.W., Piotrowski, J., Nagarajan, S., Chiotti, K., Sherlock, G., Rosenzweig, F., 2011. Hunger Artists: Yeast Adapted to Carbon Limitation Show Trade-Offs under Carbon Sufficiency. PLoS Genet 7, e1002202.

Wiedemann, B., Boles, E., 2008. Codon-Optimized Bacterial Genes Improve l-Arabinose Fermentation in Recombinant Saccharomyces cerevisiae. Appl Environ Microbiol 74, 2043–2050.

Wisselink, H.W., Toirkens, M.J., Wu, Q., Pronk, J.T., van Maris, A.J.A., 2009. Novel Evolutionary Engineering Approach for Accelerated Utilization of Glucose, Xylose, and Arabinose Mixtures by Engineered Saccharomyces cerevisiae Strains. Appl Environ Microbiol 75, 907–914.

Wright, J., Bellissimi, E., de Hulster, E., Wagner, A., Pronk, J.T., van Maris, A.J.A., 2011. Batch and continuous culture-based selection strategies for acetic acid tolerance in xylose-fermenting Saccharomyces cerevisiae. FEMS Yeast Research 11, 299–306.

Xu, Q., Singh, A., Himmel, M.E., 2009. Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. Current Opinion in Biotechnology 20, 364–371.

Yamada, R., Tanaka, T., Ogino, C., Kondo, A., 2010. Gene copy number and polyploidy on products formation in yeast. Applied Microbiology and Biotechnology 88, 849–857.

Yong, Q., Li, X., Yuan, Y., Lai, C., Zhang, N., Chu, Q., Xu, Y., Yu, S., 2012. An Improved Process of Ethanol Production from Hemicellulose: Bioconversion of Undetoxified Hemicellulosic Hydrolyzate from Steam-Exploded Corn Stover with a Domesticated <i>Pichia stipitis</i> Applied Biochemistry and Biotechnology 167, 2330–2340.

Zaldivar, J., Roca, C., Le Foll, C., Hahn-Hägerdal, B., Olsson, L., 2005. Ethanolic fermentation of acid pre-treated starch industry effluents by recombinant Saccharomyces cerevisiae strains. Bioresource Technology 96, 1670–1676.

Zhang, J., Siika-aho, M., Tenkanen, M., Viikari, L., 2011. The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. Biotechnology for Biofuels 4, 60.

Zhang, J.-G., Liu, X.-Y., He, X.-P., Guo, X.-N., Lu, Y., Zhang, B., 2011. Improvement of acetic acid tolerance and fermentation performance of <i>Saccharomyces cerevisiae</i> by disruption of the <i>FPS1</i> aquaglyceroporin gene. Biotechnology Letters 33, 277–284.

Zhang, M., Eddy, C., Deanda, K., Finkelstein, M., Picataggio, S., 1995. Metabolic Engineering of a Pentose Metabolism Pathway in Ethanologenic Zymomonas mobilis. Science 267, 240–243.

Zhang, W., Geng, A., 2012. Improved ethanol production by a xylose-fermenting recombinant yeast strain constructed through a modified genome shuffling method. Biotechnology for Biofuels 5, 46.

Zhao, X.Q., Bai, F.W., 2009. Mechanisms of yeast stress tolerance and its manipulation for efficient fuel ethanol production. Journal of Biotechnology 144, 23–30.

Zheng, D.-Q., Wu, X.-C., Tao, X.-L., Wang, P.-M., Li, P., Chi, X.-Q., Li, Y.-D., Yan, Q.-F., Zhao, Y.-H., 2011. Screening and construction of Saccharomyces cerevisiae strains with improved multi-tolerance and bioethanol fermentation performance. Bioresource Technology 102, 3020–3027.

Zhong, C., Peng, D., Ye, W., Chai, L., Qi, J., Yu, Z., Ruan, L., Sun, M., 2011. Determination of Plasmid Copy Number Reveals the Total Plasmid DNA Amount Is Greater than the Chromosomal DNA Amount in Bacillus thuringiensis YBT-1520. PLoS ONE 6, e16025.

Zhou, H., Cheng, J., Wang, B.L., Fink, G.R., Stephanopoulos, G., 2012. Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by Saccharomyces cerevisiae. Metabolic Engineering 14, 611–622.

Žifčáková, L., Baldrian, P., 2012. Fungal polysaccharide monooxygenases: new players in the decomposition of cellulose. Fungal Ecology 5, 481–489.

List of Publications

1. **Demeke MM,** Dietz H, Li Y, Foulquié-Moreno M, Mutturi S, Deprez S, Den Abt T, Bonini B , Liden G, Dumortier F, Verplaetse A, Boles E and Thevelein JM.
   "Development of a D-xylose fermenting and inhibitor tolerant industrial yeast strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering". *Biotechnology for Biofuels, 2013, **6**:89*
2. **Demeke MM,** Dumortier F, Li Y , Broeckx T, Foulquié-Moreno M and Thevelein JM. "Combining inhibitor tolerance and Dxylose fermentation in industrial Saccharomyces cerevisiae strain for efficient lignocellulose based bioethanol production". *Biotechnology for Biofuel*, 2013, **6**:120
3. Swinnen S, Schaerlaekens K, Pais T, Claesen J, Hubmann G, Yang Y, **Demeke M**, Foulquié-Moreno MR, Goovaerts A, Souvereyns K, Clement L, Dumortier F, Thevelein JM. Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. *Genome Research.* 2012 Mar 23
4. Ceyssens Pieter-Jan; Hertveldt Kirsten; Ackermann Hans-W; Noben Jean-Paul; **Demeke Mekonnen**; Volckaert Guido; Lavigne Rob. The intron-containing genome of the lytic *Pseudomonas* phage LUZ24 resembles the temperate phage PaP3. *Virology* 2008; 377(2) 233-238. (Publication from master thesis)
5. **Demeke MM**, Foulquié-Moreno M, Dumortier F, Thevelein JM. Genome analysis of the D-xylose fermenting industrial yeast strain GS1.11-26. (In preparation)

Oral presentation

1. An efficient xylose utilizing industrial *S. cerevisiae* for bio-ethanol production developed by metabolic and evolutionary engineering.
   International symposium on alcohol fuels (ISAF), 25-27 March 2013; Spier Estate, Stellenbosch, cape Town, South Africa.
2. Development of Robust Xylose Utilizing Industrial Yeast Strain for Bio-ethanol Production.
   Knowledge for growth, *FlandersBio's annual life sciences convention;* ICC Gent, Belgium; 24 May 2012.
3. Development of xylose utilizing and inhibitor tolerant industrial yeast strain for second generation bio-ethanol production.
   The International Congress on Yeast; Madisen WI, USA; 26-30 Aug 2012.
4. Development of robust xylose fermenting industrial yeast strain for bioethanol production.
   CINBIOS, a Forum for Industrial Biotechnology. 23 Oct 2012, Mechelen, Belgium.
5. Development of superior xylose-fermenting industrial yeast strains for second-generation bio-ethanol production through genome shuffling and evolutionary engineering.
   VIB seminar, January 2013. Blankenberge, Belgium.

Poster presentations

1. **Demeke MM**, Schaelaekens T, Dumortier F, Thevelein JM. Genetic mapping of mutations improving xylose fermentation in a yeast strain for bio-ethanol production.
   Recent Advances in Fermentation Technology VIII, *San Diego, California:* November 8-11, 2009.
2. **Demeke MM**, Schaelaekens T, Dumortier F, Thevelein JM. Bio-ethanol production by yeast: genetic mapping of mutations improving xylose fermentation.
   Yeasterday, Leuven, Belgium May 21st 2010

3. **Demeke MM**, Dumortier F, Foulquié-Moreno M, Thevelein JM**.** Construction of Efficient Xylose Utilizing Industrial Yeast Strain through Genome Shuffling and Evolutionary Engineering.
   VIB seminar 2012, Blankenberge, Belgium.

Press release

1. More Efficient Production of Biofuels from Waste with the Help of Modified Yeasts.
   Science Daily: Aug 29, 2013
2. VIB-vorsers verbeteren productie biobrandstoffen uit afval
   Knack: 29/08/2013